UNIVERSITÀ DEGLI STUDI DI PALERMO
Dottorato di Ricerca in Ingegneria dell'Innovazione Tecnologica
Dipartimento di Ingegneria
ING-INF/05

IMAGE PROCESSING, SEGMENTATION AND MACHINE LEARNING MODELS TO CLASSIFY AND DELINEATE TUMOR VOLUMES TO SUPPORT MEDICAL DECISION

IL DOTTORE
Dott. Albert Comelli

IL COORDINATORE
Prof. Salvatore Gaglio

IL TUTOR
Prof. Giovanni Petrucci

CO TUTOR
Prof. Anthony Yezzi

CICLO XXXII
ANNO CONSEGUIMENTO TITOLO 2020

ABSTRACT

Techniques for processing and analysing images and medical data have become the main's translational applications and researches in clinical and pre-clinical environments. The advantages of these techniques are the improvement of diagnosis accuracy and the assessment of treatment response by means of quantitative bio-markers in an efficient way. In the era of the personalized medicine, an early and efficacy prediction of therapy response in patients is still a critical issue.

In radiation therapy planning, Magnetic Resonance Imaging (MRI) provides high quality detailed images and excellent soft-tissue contrast, while Computerized Tomography (CT) images provides attenuation maps and very good hard-tissue contrast. In this context, Positron Emission Tomography (PET) is a non-invasive imaging technique which has the advantage, over morphological imaging techniques, of providing functional information about the patient's disease.

In the last few years, several criteria to assess therapy response in oncological patients have been proposed, ranging from anatomical to functional assessments. Changes in tumour size are not necessarily correlated with changes in tumour viability and outcome. In addition, morphological changes resulting from therapy occur slower than functional changes. Inclusion of PET images in radiotherapy protocols is desirable because it is predictive of treatment response and provides crucial information to accurately target the oncological lesion and to escalate the radiation dose without increasing normal tissue injury. For this reason, PET may be used for improving the Planning Treatment Volume (PTV). Nevertheless, due to the nature of PET images (low spatial resolution, high noise and weak boundary), metabolic image processing is a critical task.

The aim of this Ph.D thesis is to develope smart methodologies applied to the medical imaging field to analyse different kind of problematic related to medical images and data analysis, working closely to radiologist physicians.

Various issues in clinical environment have been addressed and a certain amount of improvements has been produced in various fields, such as organs and tissues segmentation and classification to delineate tumors volume using meshing learning techniques to support medical decision.

In particular, the following topics have been object of this study:

- Technique for Crohn's Disease Classification using Kernel Support Vector Machine Based;
- Automatic Multi-Seed Detection For MR Breast Image Segmentation;
- Tissue Classification in PET Oncological Studies;
- KSVM-Based System for the Definition, Validation and Identification of the Incisinal Hernia Reccurence Risk Factors;
- A smart and operator independent system to delineate tumours in Positron Emission Tomography scans;

- Active Contour Algorithm with Discriminant Analysis for Delineating Tumors in Positron Emission Tomography;
- K-Nearest Neighbor driving Active Contours to Delineate Biological Tumor Volumes;
- Tissue Classification to Support Local Active Delineation of Brain Tumors;
- A fully automatic system of Positron Emission Tomography Study segmentation.

This work has been developed in collaboration with the medical staff and colleagues at the:

- Dipartimento di Biopatologia e Biotecnologie Mediche e Forensi (DIBIMED), University of Palermo
- Cannizzaro Hospital of Catania
- Istituto di Bioimmagini e Fisiologia Molecolare (IBFM) Centro Nazionale delle Ricerche (CNR) of Cefalù
- School of Electrical and Computer Engineering at Georgia Institute of Technology

The proposed contributions have produced scientific publications in indexed computer science and medical journals and conferences. They are very useful in terms of PET and MRI image segmentation and may be used daily as a Medical Decision Support Systems to enhance the current methodology performed by healthcare operators in radiotherapy treatments.

The future developments of this research concern the integration of data acquired by image analysis with the managing and processing of big data coming from a wide kind of heterogeneous sources.

ABBREVIATIONS

BTV: Biological Tumour Volume

CT: Computerized Tomography

DICOM: Digital Imaging and Communications in Medicine

FBP: Filtered Back-Projection

FDG: 8F-fluoro-2-deoxy-D-glucose

GTV: Gross Tumour Volume

HNC: Head and Neck Cancer

MET: 11C-labeled Methionine

MRI: Magnetic Resonance Imaging

NSCLC: Non-Small Cell Lung Cancer

PET: Positron Emission Tomography

PVE: Partial Volume Effect

PTV: Planning Treatment Volume

ROI: Region Of Interest

SUV: Standardized Uptake Value

TPS: Treatment Planning Systems

Contents

CHAPTER 1

1 Introduction

The medical imaging is a crucial field for diagnostic and treatment purposes. Several medical imaging techniques are available and their improvement is one of the most important goals in the health's research.

1.1 Medical Imaging

Medical imaging is characterized by great multitude of heterogeneous data enclosing several experts and expertise in different fields, such as physicians, engineers, biologists, and physics for disease staging and treatment purposes.

According to the source and the physical properties, the medical imaging techniques can be differentiated in two groups:

- anatomical/structural images, and
- functional images.

In particular, Positron Emission Tomography (PET), Computerized Tomography (CT), and Magnetic Resonance Imaging (MRI) are *in-vivo* tomographic techniques that allows showing axial, sagittal and coronal images without moving the patient. Furthermore, these techniques can be classified on the use of ionized or non-ionized radiation as well as external or internal sources.

To support clinical decisions is mandatory to develop computer-assisted methods of quantitative biomarker extraction by medical images. To date, the computational power is mature: an effort is mandatory to address this power to analyze and process medical data in the best possible way. Fast and operator independent approaches are the crucial keywords in clinical environment to obtain an effective impact in the work of medical operators involved in diagnosis and treatment assessment.

Developing computer decision support systems that integrate the knowledge of the medical experts, and facilitate the diagnostic procedure are need for a better patient management. In particular, computer-assisted methods for segmenting regions of interests (ROIs) in medical images are increasingly important in assisting and automating specific radiotherapy tasks.

1.1.1 DICOM Protocol

The medical images produced by special diagnostic equipment, such as PET, CT, and MRI, are compliant with the DICOM (Digital Imaging and Communications in Medicine) standard.

DICOM standard was created by National Electrical Manufacturers Association

(NEMA) and it enables users to retrieve images and related information from different medical imaging modalities with a standardized way allowing the communication between electronic devices and biomedical computing from different vendors facilitating the management of digital images. In this way, the introduction of new services that support the medical applications is facilitated: DICOM enables the integration of scanners, servers, workstations, printers, and network hardware from multiple manufacturers.

A DICOM file consists of a header (so called metadata) and image data. The header includes image related information such as image type, study, information on the mode of acquisition, resolution (height and width), colour, size of the voxel, number of stored bits, and patient's information.

In conclusion, the DICOM standard describes a directory structure that facilitates access to medical images and related information.

1.1.2 PET Imaging

PET is a non-invasive nuclear medical imaging technique based on the visualization of functional processes showing complementary information with respect to anatomical imaging. It provides an in vivo measure of the tumor biological processes [1]. Among several PET radiotracers derived from isotopes, 18F-fluoro-2-deoxy-D-glucose (FDG) is a glucose analogue widely used in the evaluation of several neoplastic pathologies as well as in radiotherapy planning. FDG uptake is increased in tissue with a high metabolic rate, such as tumor or inflammation regions. These areas appear as hot spots on PET images. FDG PET is able to identify the location of many primary tumors and metastases offering the opportunity to radically change patient treatment (i.e. from radiotherapy to chemotherapy) or the radiotherapy planning treatment volume (PTV) [2]. For these reasons, the oncology field is the most used and developed for PET studies: this technique has been recognized as an adequate staging and restaging tool in various cancer types. In addition, metabolic changes are often faster and more indicative of the effects of the therapy with respect to morphological changes [3]. To obtain a tomographic study, many acquisitions of the same object at different angles are executed; the object is reconstructed in three dimensions using complex mathematical algorithms such as the Filtered Back-Projection algorithm (FBP) or iterative methods. After applying reconstruction algorithms, a digital image that represents the uptake distribution in the tissues is obtained. Figure 1.1.1 an example of PET study (coronal, sagittal, and axial imaging planes).

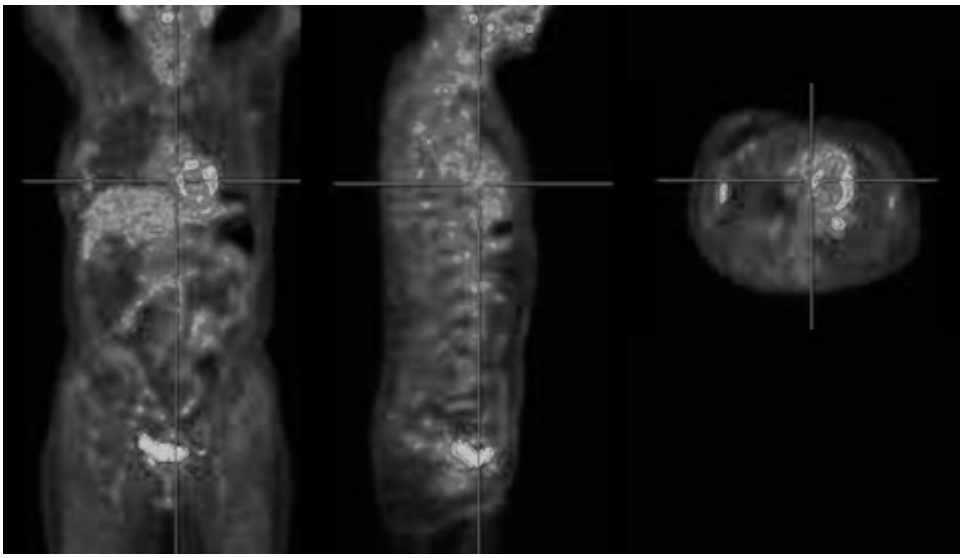


Figure 1.1.1 An example of PET study (coronal, sagittal, and axial imaging planes).

1.1.3 CT Imaging

CT is a diagnostic imaging technique that uses ionizing radiation. CT studies are acquired by rotating an X-ray source around the patient. X-ray sensors are positioned on the opposite side, and the patient's bed is continuously moved to obtain a helical acquisition mode. X-ray beam intensity exponentially decreases in extent to the mass attenuation coefficient: the beam will be more attenuated by tissues with a high atomic number. Vice versa, if the beam crosses a low density tissue, the attenuation will be less. In this way, high-density tissues appear clear (maximum attenuation) and lower-density tissue appear dark (minimum attenuation).

After reconstruction, the CT images reflect the attenuation of each voxel according to the Hounsfield scale, where air has a radio-density of -1000 Hounsfield Units (HU) and distilled water a radio-density of 0 HU. CT provides high resolution morphological images of the body, with an excellent contrast between air, adipose tissue, soft tissue and bone. However, the soft tissue contrast is poor if compared to other imaging modalities such as MRI. Contrast agents may additionally be used to highlight some regions, such as the gastrointestinal tract or the blood vessels. Figure 1.1.2 an example of CT acquisition (coronal, sagittal, and axial imaging planes).

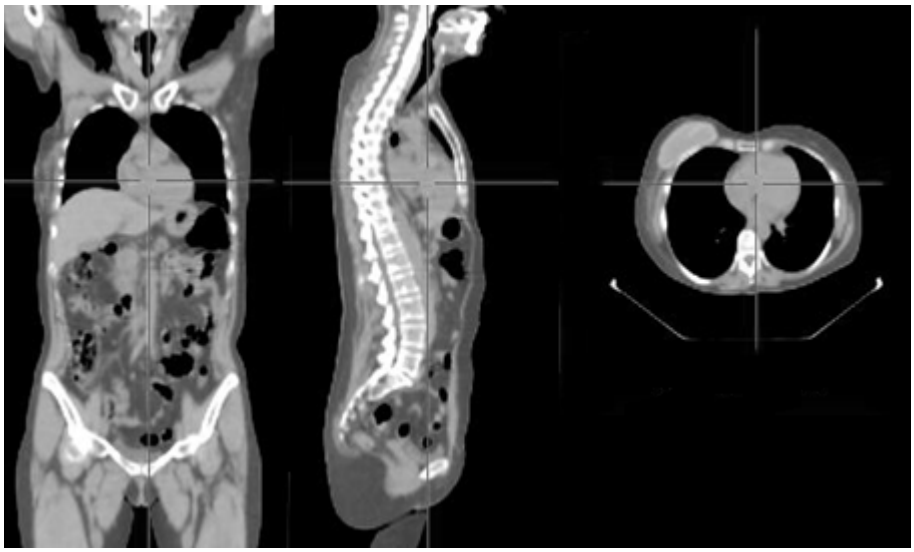


Figure 1.1.2 An example of CT acquisition (coronal, sagittal, and axial imaging planes).

1.1.4 MR Imaging

MRI is based on the different relaxation times of tissues after being subjected to an electromagnetic stimulus:

- *T1-weighted images*: the time T1, or longitudinal relaxation time is a measure of the time for which the protons to return to the initial equilibrium conditions, through the transfer of energy to the surrounding microenvironment (lattice), in order to obtain a T1-weighted SE sequence, using a short relaxation time (TR) associated with a short echo time (TE). On T1-weighted images, the cerebrospinal fluid is dark while the fat is brilliant.
- *T2-weighted images*: the time T2, or transverse relaxation time, is a measure of the time taken by the spin of protons to get out of sync. This progressive desynchronization will void the transverse magnetization. A sequence to get a T2-weighted sequence will have a long TR associated with a long TE. Liquids or at least very hydrated tissues, appear bright white in T2-weighted images.

The relaxation time of a tissue is related with its water content.

MRI provide excellent soft tissue contrast. For this reason, MRI is widely applied in the diagnosis and treatment of neurological, cardiovascular, musculoskeletal, liver and gastrointestinal diseases. Moreover, contrast can be further enhanced through the injection of a contrast enhancement agent.

1.2 Radiation Therapy Planning

Radiotherapy aims to deliver the necessary therapeutic dose of ionizing radiation to the oncological lesions minimizing irradiation to normal tissues. In this way, radiation is delivered to target to damage and to kill the cancerous cells stopping them from regenerating. Since both cancer and healthy tissues are affected by radiation, any treatment plan should be designed in such a way that the radiation dose delivered to the tumor is high enough to destroy the cancer cells avoiding the delivery of excessive doses of radiation to surrounding normal tissue. Precise tumour volume delineation is a very critical step in order to ensure safe and effective radiation therapies.

The Radiation Therapy Planning (RTP) is to implement the treatment strategy assigning the desired dose to the various ROI. Recent advances in radiotherapy, such as intensity modulated radiation therapy, have improved the dose painting of RTP. These techniques enable a precise delivery of a high dose to the target maintaining a low radiation dose to nearby critical organs. However, the hardware precision in the delivering radiation dose is far greater than the software precision in the target volume delineation: accurate target volume definition is essential for escalating the radiation dose without increasing normal tissue injury.

CT and MRI are considered the standard for target volume delineation in many cancer district. Radiotherapy results based on morphological target volume delineation are disappointing due to the radio resistance of the tumor and/or inadequate dose to target due to missed lesion. CT or MRI imaging may not show the viable extension of tumors and not localize isolated positive lymph nodes [5]. To improve these results, PET has been introduced in the radiotherapy field to assist the radiation oncologist in RTP: within the Gross Tumor Volume (GTV), defined on anatomical images, it is possible to define target volumes based on functional area (BTV – Biological Target Volume) and to apply a strategy that will deliver radiation to these regions. Though the use of PET images has been shown to improve target volume definition by reducing intra and inter-observer variability compared to CT images only [6].

1.3 Thesis Outline

The thesis is structured as follows:

- Chapter 2 describes the state-of-the-art of segmentation approaches in medical imaging.
- Chapter 3 describes an innovative method to segmentation MR breast Image.
- Chapter 4 presents Machine Learning and Data Analysis system in clinical applications to predict the disease in patients.
- Chapter 5 describes new segmentation methods for tumors delineation in PET.

- Chapter 6 shows the experimental results.
- Discussion and conclusion about this thesis.

1.4 Research Contributions

1. **A.Comelli**, M. C. Terranova, L. Scopelliti, S. Salerno, F. Midiri, G. Lo Re, G. Petrucci, and S. Vitabile. A Kernel Support Vector Machine Based Technique for Crohn's Disease Classification in Human Patients. CISIS 2017: In book: Complex, Intelligent, and Software Intensive Systems, pp.262-273. DOI: 10.1007/978-3-319-61566-0_25
2. **A.Comelli**, A. Bruno, M. L. Di Vittorio, F. Ienzi, R. Lagalla, S. Vitabile, and E. Ardizzone. Automatic Multi-Seed Detection For MR Breast Image Segmentation. Conference: International Conference on Image Analysis and Processing (ICIAP) 2017, At Catania (Italy), Volume: Vol LNCS 10484, pp. 683-693. DOI: 10.1007/978-3-319-68560-1_63
3. **A.Comelli**, A. Stefano, V. Benfante, and G. Russo. Normal and Abnormal Tissue Classification in PET Oncological Studies. Pattern Recognition and Image Analysis, 2018, Vol. 28, No. 1, pp. 121–128. DOI: 10.1134/S1054661818010054. Impact Factor: 0.998.
4. M.C.Terranova, **A.Comelli**, L.Scopelliti, F.Vernuccio, D.Picone, A.Di Piazza, C.Tudisca, F.Midiri, S.Salerno, R.Lagalla, S.Vitabile and G.Lo Re. A New Semi-Automatic Technique for Crohn's Disease Diagnosis Using Supervised Machine Learning Algorithm: Kernel Support Vector Machine. Radiological Society of North America (RSNA) 2017.
5. G.Russo, **A.Comelli**, A. Stefano, M.G.Sabini, M.Ippolito, M.C.Gilardi, A.Yezzi. Sviluppo di un metodo per la segmentazione del BTV accurato e operatore indipendente. 10°Congresso Nazionale (AIFM) 2018.
6. L.Licari, G.Salamone, S.Campanella, F.Carfì, T.Fontana, N.Falco, R.Tutino, P.De Marco, **A.Comelli**, D.Cerniglia, G.Petrucci, S.Vitabile and G.Gulotta. Use of the KSVM-Based System for the Definition, Validation and Identification of the Incisinal Hernia Reccurence Risk Factors. Journal of Surgery 2018. Impact Factor 0.95.
7. **A.Comelli**, A. Stefano, G. Russo, M.G. Sabini, M. Ippolito, S. Bignardi, G. Petrucci, A. Yezzi. A smart and operator independent system to delineate tumours in Positron Emission Tomography scans, Computers in Biology and Medicine. Accepted 6 September 2018. DOI: 10.1016/j.compbiomed.2018.09.002. Impact Factor: 2.115.
8. **A.Comelli**, A.Stefano, S.Bignardi , G.Russo, M.G.Sabini, M.Ippolito, S.Barone and A.Yezzi. Active Contour Algorithm with Discriminant Analysis for Delineating Tumors in Positron Emission Tomography. Artificial Intelligence in Medicine. Accepted 7 January 2019. DOI: 10.1016/j.artmed.2019.01.002. Impact Factor: 2.879.

9. G.Russo, **A.Comelli**, A.Stefano, M.G.Sabini, M.Ippolito, M.C.Gilardi, & Yezzi. 253. An accurate and operator independent method for biological tumour volume segmentation. Physica Medica: European Journal of Medical Physics, 56, 218. Accepted December 2018. DOI: 10.1016/j.ejmp.2018.04.264.
10. **A.Comelli**, A.Stefano, G.Russo, S.Bignardi, M.G.Sabini, G.Petrucci, M.Ippolito, and A.Yezzi. K-Nearest Neighbor driving Active Contours to Delineate Biological Tumor Volumes. Engineering Applications of Artificial Intelligence. Under review 2018. DOI: 10.1016/j.engappai.2019.02.005.
11. **A.Comelli**, A.Stefano, S.Bignardi, C.Coronnello, G.Russo, M.G.Sabini, M.Ippolito, and A.Yezzi. Tissue Classification to Support Local Active Delineation of Brain Tumors. In Annual Conference on Medical Image Understanding and Analysis. Springer, Cham, 2019.
12. **A.Comelli**, and A.Stefano. A fully automatic system of Positron Emission Tomography Study segmentation. In Annual Conference on Medical Image Understanding and Analysis. Springer, Cham, 2019.

CHAPTER 2

2 Review of the State of the Art

2.1 Medical Image Segmentation

Image segmentation is a fundamental and critical task in numerous biomedical imaging applications. It corresponds to extract a target out of an image identifying edges and areas of similar features.

Various delineation methods have been proposed in the literature. Each segmentation method has a different approach to find the optimal segmentation: depending on the application, appropriate segmentation algorithm must be chosen. It is impossible to make a categorical statement; different segmentation methods have different goals, and each method may be useful in different imaging applications.

In this study, the aim is to segment functional and anatomical regions for RTP. In the next paragraphs, various methods for BTV and multimodality image delineation are discussed.

2.1.1 PET, MRI, and CT Segmentation

Most of the segmentation techniques on medical images have been applied on CT, MR or ultrasound images and not on PET images since PET has received increasing attention only in recent years in order to incorporate metabolic information in RTP and to evaluate early treatment response in oncological patients.

To date, clinical analyses of PET studies are performed on software platforms with predefined window-level setting. The nuclear physician, altering the visual appearance of PET lesions, can liberally modify this setting. The manual segmentation approach is then dependent on the experience of the nuclear physician limiting the measurement accuracy. In addition, the manual segmentation is time-consuming and impaired by inter and intra-observer variability.

Currently available tumor delineation methods in PET imaging are based on a fixed threshold of the maximum tracer uptake value in the lesion. Other methods based on fuzzy c-means (FCM) [7], region growing [8] or watershed segmentation [9] have been suggested, but few validation studies are available and there is no consensus for a proper BTV delineation method with no clear guidelines on how to incorporate metabolic data into target delineation [10]. Moreover, PET delineation approaches can be categorized on the basis of anatomical sites or radio-tracers used

and, according to a comprehensive review of segmentation algorithms in PET imaging [11], some segmentation methodology categories can be identified.

Image threshold methods are the most widely used due to their intuitive basis and simple implementation. Usually, each PET image voxel is initially converted into its corresponding body-weight Standardized Uptake Value (SUV). The SUV is a widely used PET semi-quantitative parameter, calculated as a ratio of tissue radioactivity concentration and FDG injected dose at the time of injection divided by body weight. Subsequently, a fixed, or adaptive, or iterative threshold [12][13][14][15] based on phantom simulations and on scanner hardware features is calculated. An absolute 2.5 SUV value, independent of background and lesion volume, was proposed by Paulino et al. [16] to discriminate benign from malignant lesions and to delineate the BTV. This approach is very controversial and not accepted by the whole scientific community. However, some fixed threshold methods based on analytical equations extracted from realistic phantom experiments can have a better accuracy with respect to more complex adaptive or iterative approaches. These techniques need to be calibrated for each PET scanner and should not be used without optimization. Generally, image threshold methods are inter- and intra-operator independent but strongly scanner dependent, requiring a significant effort in the calibration data to identify the equation parameters. In addition, they are too sensitive to image noise and heterogeneity and partial volume effect (PVE) due to the limited spatial resolution of the PET scanner [17].

Region growing is a recursive algorithm that groups pixels or sub-regions into larger regions. The grouping is based on predefined criteria, such as grey level threshold values. A region growing usually starts off with one seed point and then recursively add neighbouring pixels that fulfil the criteria to the region. Different connectivities can be used, in 2-D region growing the choice is usually between four and eight-connectivity. The algorithm ends by itself when no more pixels fulfil the criteria on a pixel level [8].

Supervised and unsupervised learning methods such as artificial neural networks (ANN), support vector machines (SVM), k-means algorithms, FCM algorithms are efficient but only in large lesions of simple shape. These approaches discriminate target from background based on a set of image features. However, supervised algorithms (ANN and SVM) have limited application in PET imaging, unlike in the MRI or CT fields, due to high heterogeneity that makes the recognition of stable features in the training set difficult. On the other hand, clustering methods such as FCM or Fuzzy Locally Adaptive Bayesian are optimal for the fuzzy nature of the PET lesion edge where a finite number of levels, rather than only two (foreground and background), is used to label voxels within the target area [7][17]. The main issue is the correct identification of the cluster number; Zaidi et al. [7] started the algorithm with an oversized number of clusters followed by a merging process to group clusters with similar properties to reach a number of clusters in agreement with uniform attenuation areas based on anatomical delineation. However, learning

methods require high computational steps, especially in supervised methods, and they are sensitive to image heterogeneity and to study protocol variability, as for example scanner characteristics, PET radiotracers, radiotracer injected dose and interval between radiotracer injection and examination start.

Variational approaches are based on gradient differences between target and background regions attempting to exploit information provided by intensity variation [18]. They require smooth regions with reliable boundary information but PET images are characterized by low contrast. In the study proposed by El Naqa et al. [19], deformable active contour models are computed as directional gradients by means of energy minimization. These approaches are mathematically efficient but require an initialization and are sensitive to image noise and subject to numerical fluctuation.

In clinical routine, it is highly desirable to have both functional and structural quantifiable information so the disease can be both identified and localized, potentially resulting in an earlier diagnosis and more effective treatment plan. A concise but complete review of the state of the art regarding multimodal co-segmentation approaches is reported in the following.

Several studies approach PET/CT tumor identification and characterization in radiation therapy scenarios. In [20], co-registered FDG-PET/CT were used for the textural characterization of head and neck cancer (HNC) for radiotherapy treatment planning. After a manual segmentation on co-registered PET/CT images (performed by an experienced radiation oncologist), useful textural features were selected for distinguishing tumor from normal tissue in HNC subjects. Both k-nearest neighbours (KNNs) and decision tree (DT)-based KNN classifiers were employed to discriminate images of cancerous and healthy tissues. Han et al. [21] presented a Markov random field (MRF)-based co-segmentation of the PET/CT image pair with a regularized term that penalizes the segmentation difference between PET and CT. This graph-based method utilizes the strength of PET and CT modalities for target delineation in a group of 16 patients with HNC. Background and foreground seed voxels must be always manually identified by the user. A similar approach is reported in [22], where the segmentation is seen as a minimization problem of a MRF model, which encodes the information from both modalities. This optimization is solved using a graph-cut based method, by constructing two sub-graphs for PET and CT segmentation, respectively. The algorithm was validated in robust delineation of lung tumors on 23 PET/CT datasets and two HNC subjects. A further MRF-based systematic solution for the automated co-segmentation of brain PET/CT images into grey matter, white matter and CSF regions is exposed in [23]. A PET/CT image pair and its segmentation result are modelled as a MRF triplet, and segmentation is eventually achieved by solving a maximum a posteriori (MAP) problem using the expectation maximization (EM) algorithm with simulated annealing. The overall MRF-MAP model was tested both on both simulated and real patient PET/CT data.

The authors of [24] proposed a method for automated delineation of tumor boundaries in whole-body PET/CT by jointly using information from both PET and diagnostic CT images. After an initial robust hot-spot detection and segmentation performed in PET, a model for tumor appearance and shape in corresponding CT structures is learned by weighted non-parametric density estimate. This voxel-based CT classification is then probabilistically integrated with PET classification using the joint likelihood ratio test technique to derive the final segmentation. The algorithm was tested on patient studies with lung and liver tumors identifiable in both the PET and CT images acquired using the same scanner.

Yezzi et al. [25] introduced a geometric variational framework that uses active contours to simultaneously segment and register features from multiple images. The key aspect of this approach is that multiple images may be segmented by evolving a single contour as well as the mappings of that contour into each image during feature-based realignment steps. The results of three experiments on MRI/CT images of the head and the spine are reported. Also the authors of [19] developed variational methods based on multivalued level set deformable models for simultaneous 2D or 3D segmentation of multimodality images consisting of combinations of co-registered PET, CT, or MRI datasets. In particular, only three patients are considering: a non-small cell lung cancer case with PET/CT, a cervix cancer case with PET/CT, and a prostate patient case with CT/MRI. In addition, CT, PET, and MRI phantom data were used for quantitative validation of the proposed multimodality segmentation approach.

An automatic algorithm for the co-segmentation of HNC based on PET/MRI data was proposed in [26], in order to standardize tumor volume delineation. For both imaging modalities tumor probability maps were derived, assigning each voxel a probability of being cancerous according to its signal intensity. A combination of these maps was subsequently segmented using a threshold level set algorithm. The algorithm processes both the anatomical T2-weighted MRI and FDG-PET data concerning 10 HNC patient datasets acquired in a combined PET/MRI system.

The group led by Bagci developed co-segmentation approach in multimodal medical imaging, using the Random Walker (RW) algorithm [27] and unifying graph representation of each image modality in a single product lattice. The overall method results in a fully automatic framework, providing an automated object detection via interesting uptake region algorithm to avoid users foreground and background seed detection. Afterwards, prior to the initiation of the segmentation process, these identified seeds are propagated to the corresponding anatomical images. Although no significant anatomical and functional changes between the scans have to be assumed, the study used PET, PET/CT, MRI/PET, and fused MRI/PET/CT scans from 56 patients who had various lesions in different body regions.

CHAPTER 3

3 MR Breast Image Segmentation

In breast MRI, several elements are required to perform automatic analysis. Many examples of medical imaging require an initial segmentation phase: multimodal breast image registration, computer aided analysis of DCE (dynamic contrast enhanced) MRI [30], and breast density assessment [1], [29]. In [30] the authors detected the left side and the right side of the breast and the center of mass in each side is used as the seed points for region growing. The region detection and extraction from the anatomical regions are very difficult tasks. Complicating factors are the large shape variations of pectoral muscles across different patients, the similarity between intensity distributions and texture descriptors of the breast MR in muscle and fibroglandular tissues.

In last few years, many researches appears in medical imaging and precise segmentations of relevant anatomical structures such as breast region and fibroglandular tissue are required. Most of state of the art methods for breast segmentation on MRI are semi or fully automated, furthermore they can be grouped in contour-based, region-based and atlas-based approaches [31]. Generally, on Breast MRI, the following operations precede the breast segmentation task: Pectoralis muscle boundary segmentation, breast-air boundary segmentation. In [32] the authors proposed a method based on the observation that the pectoralis muscle and breast-air boundaries exhibit smooth sheetlike surfaces in 3D. This surfaces that can be simultaneously enhanced by a Hessian-based sheetness filter. The authors in [33] proposed a method for breast segmentation, but it needs manual intervention. In [34] breast segmentation was based on a semiautomated model that accounting for partial volume effects.

3.1 Automatic Multi-seed Detection for MR Breast Image Segmentation

The proposed method consists of three steps: (1) pre-processing step to locate three regions of interest (axillary and sternal regions); (2) processing step to detect maximum concavity points for each region of interest; (3) breast image segmentation step. Eighteen patients have been manually segmented accordingly to three expert Radiologists to generate Gold Standard ground-truth used to evaluate the effectiveness of the proposed method. The acquisition parameters and characteristics are depicted in the next section. The algorithms used in the proposed segmentation method are briefly described in the next sections. The proposed system consists of three main steps, as depicted in Figure 3.1.:

- The pre-processing output as depicted in Figure 3.1.1;
- The processing output as depicted in **Figure 3.1.2**;
- The segmentation output as depicted in **Figure 3.1.3**.

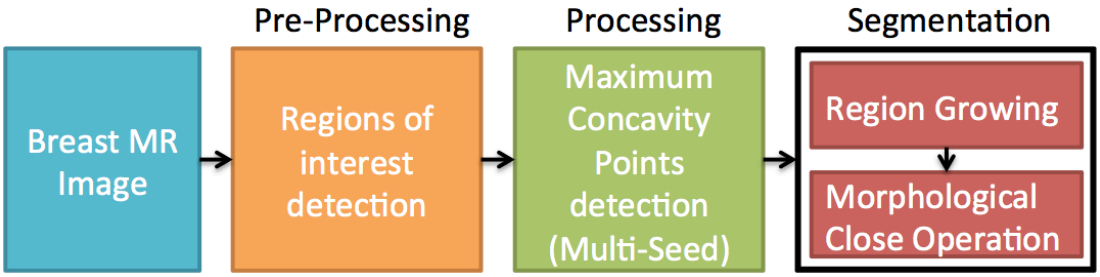


Figure 3.1.1 Block scheme of the proposed method: pre-processing, processing, and segmentation.

3.1.1 Pre-processing

The pre-processing step is as follows:

1. A breast MRI study is loaded. It contains, more or less 30 slices with thickness 5.00mm (a single slice as shown in **Figure 3.1.1** a);
2. A first binarization step is required to extract the boundary of the breast as shown in **Figure 3.1.1** b), an adaptive thresholding is applied to the image by analyzing the trimodal distribution of intensity histogram;
3. An image crop containing the breast regions is considered (approximatively 2/3 of the whole image) as shown in **Figure 3.1.1** c);
4. The holes are filled in the image, as shown in **Figure 3.1.1** d);
5. The Largest Connected Component is found in the MRI and all other components are removed, as shown in **Figure 3.1.1** e);
6. The objective is to find the coordinates of three pairs of points (A, B, C, D, E, F as shown in **Figure 3.1.1** f), each pair of points includes a concave region of the breast boundary. The first and the third pair of points correspond to the vertices of axillary regions; the second pair of points correspond to the sternalregion. The task is to find an n-by-2 matrix that specifies the convex hull includingthe Breast Region and each row includes the coordinates of the convex hull corners. We notice that the larger side of the convex hull correspond to our regions of interest i.e. the axillary and stern regions (as suggested by the radiologists). In few words we find the three pairs of points by detecting and sorting the larger sides of the aforementioned convex hull. The first three sides of the convex polygon respectively correspond to the axillary (A, B, E, F see in **Figure 3.1.1** f) and the sternal regions (C, D see in **Figure 3.1.1** f). We sort the vector including the distances between the consecutive vertices of the convex polygon, in descend order, then we select the first three pairs of coordinates as the vertices of our regions of interest (axillary and sternal regions), as shown in **Figure 3.1.1** f;

7. The convex hull of the binary regions found in step 5 is computed, as showing **Figure 3.1.1** g;
8. The boundary of image obtained in item 5 is extracted with the canny filter, as shown in **Figure 3.1.1** h;

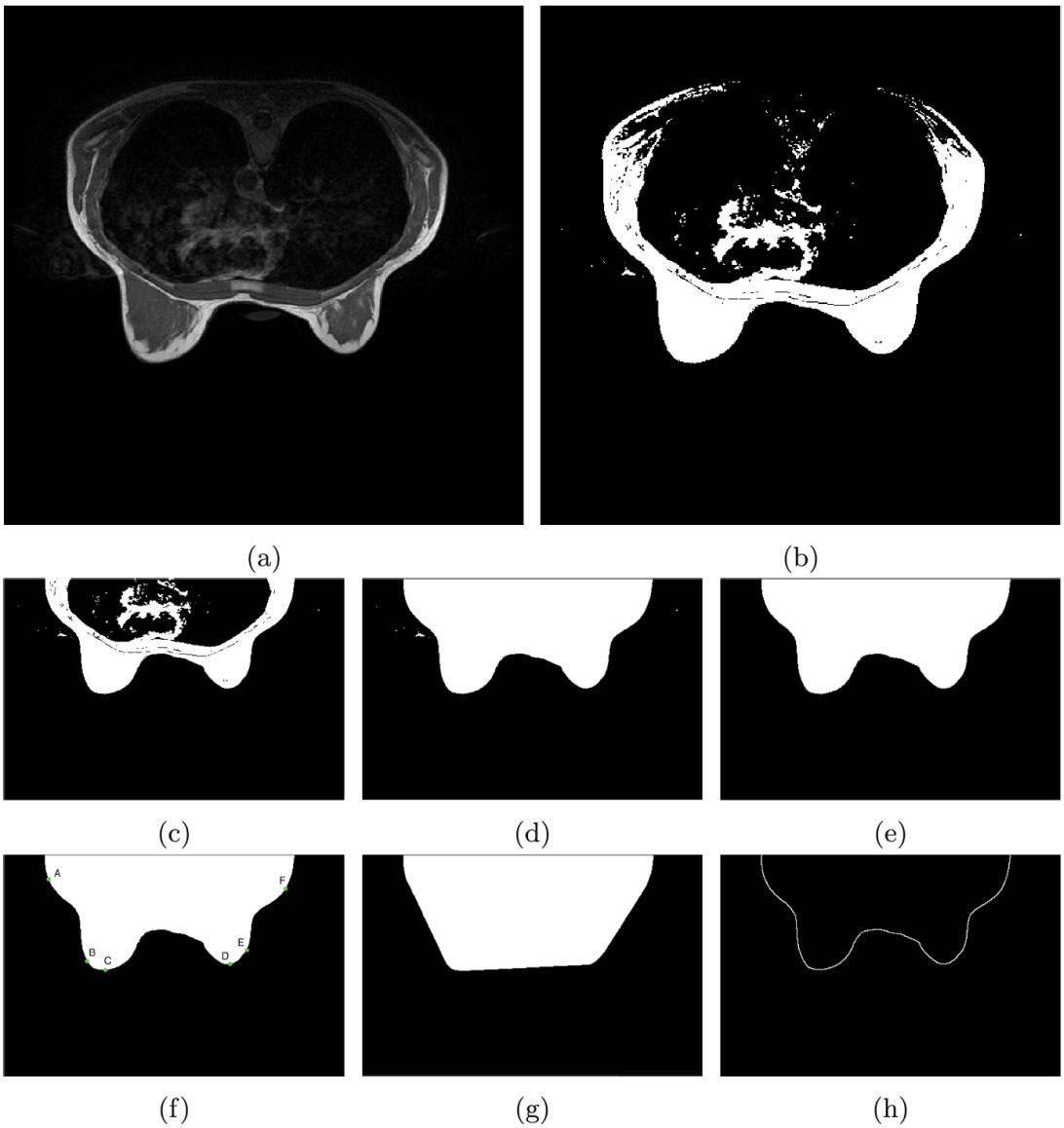


Figure 3.1.1 Pre-processing: (a) a breast MRI is loaded; (b) the adaptive thresholding is applied in the MRI to emphasize the contours of the breast; (c) image crop containingthe breast regions is considered; (d) the holes are filled in the image obtained in (c); (e) the Largest Connected Component is found in the MRI and all other components are removed; (f) the coordinates of three pairs of green points (A, B, C, D, E, F) are found; (g) the convex hull is computed in the image obtained in Figure e and it is returned a binary convex hull image; (h) the boundary of image obtained in Figure e is extracted with the canny filter.

3.1.2 Processing

The coordinates of the points A–F (see in Figure 3.1.1 f) are grouped in three pairs:

A and B belong to the first side of the convex hull (axillary region), C and D belong to the second side of the convex hull (stern region), E and F belong to the third side of the convex hull (axillary region). To detect the maximum concavity points (in breast image, see green points G, H, I in Figure 3.1.2 c) we process three regions of interest i.e. the axillary regions and the stern region. We highlight that the maximum concavity points correspond to the landmarks identified by the radiologists. To accomplish the detection of the maximum concavity points, each region of interest is processed as it follows:

1. To detect the maximum concavity, we first apply Delaunay Triangulation between the two vertices of each region of interest (in first instance A–B in **Figure 3.1.1** f) and all the boundary points between the two vertices (yellow dots in **Figure 3.1.2** a);
2. The area of each triangles is computed;
3. The triangles are sort in descend order with respect to area value;
4. The first triangle is selected, it includes the larger area in the concave region of interest;
5. The maximum concavity point of the region is the third vertex of the triangle selected in the previous step (see green point G in **Figure 3.1.2** c);
6. To avoid the exclusion of some region of interest including important features such as Skin, subcutaneous fat pad, and chest fat pad we add and additional margin to the G coordinates;
7. The additional margin is computed by measuring the vertical distance between G and the inner boundary extracted by filtering with canny algorithm. The same technique is applied to the other maximum concavity points, as depicted in **Figure 3.1.2** c) by red points;
8. When the maximum concavity points are detected, a line is drawn to join these points. All the elements located above the line are deleted, as shown in **Figure 3.1.2** d).

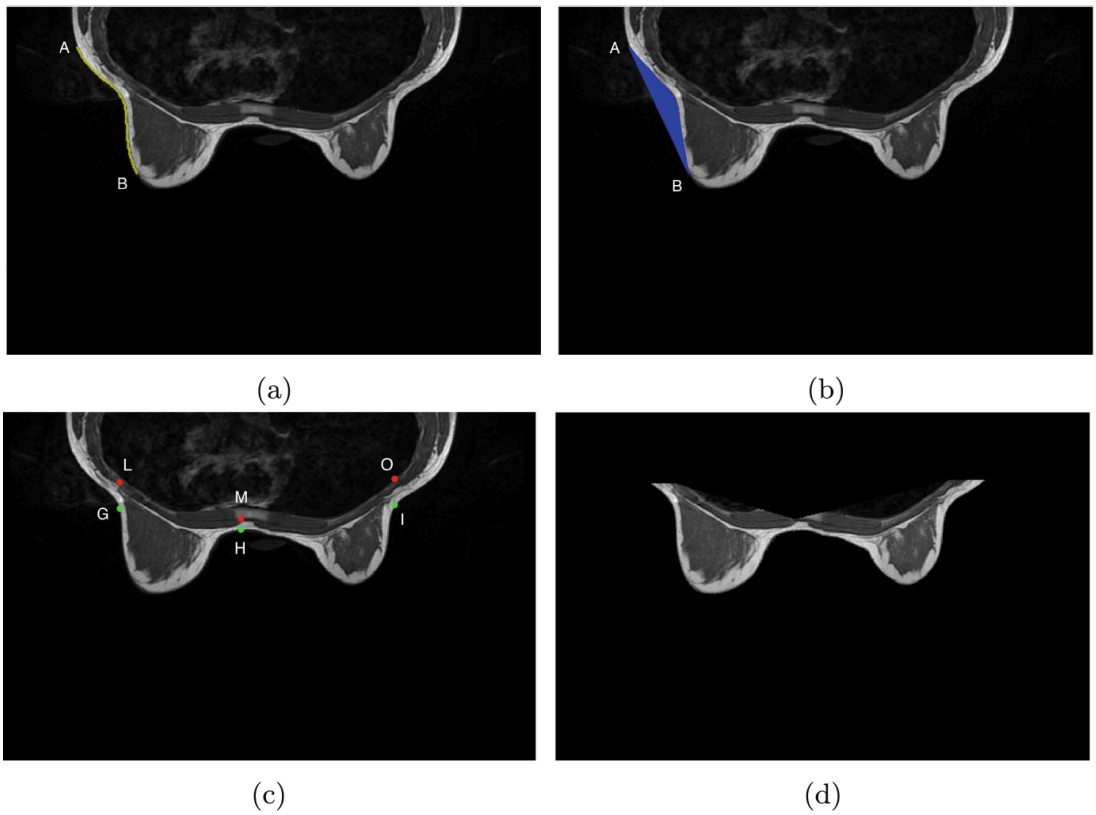
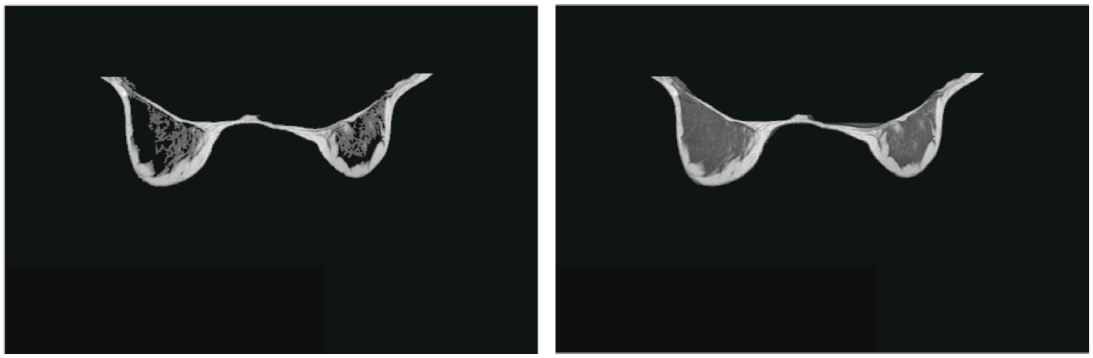


Figure 3.1.2 Processing: (a) the boundary points between the two vertices A–B (yellow dots) are plotted; (b) Delaunay Triangulation is applied between the two vertices of each region of interest (in first instance A–B) and all the boundary points between the two vertices (yellow dots in (a); (c) for each concavity to detect the maximum concavity points (in breast image, see green points G, H, I); an additional margin is computed by measuring the vertical distance between G and the inner boundary extracted by filtering with canny algorithm, as depicted by red points in (c); (d) when the extraction of concavity points stage is complete, a line is drawn to join these points. All components above this line are removed.

3.1.3 Segmentation

The segmentation phase consists of three steps:

1. First, a region growing algorithm [37] with standard parameters (threshold) is applied to the image processed as described in the previous section. Region Growing algorithm needs a seed point to be executed. The maximum concavity points are then used as seed points for region growing. The result is shown in **Figure 3.1.3** a);
2. In second step, morphological close operation has been used to fill the holes emerged from region growing. The structuring element of morphological operations is a disk with radius of 20 pixels so that the largest hole gets filled. The disk structuring element is used to preserve the circular nature of the object. The result is shown in **Figure 3.1.3** b);



(a) (b)

Figure 3.1.3 Segmentation: (a) the region growing is applied to the image obtained in the end of processing step Fig. 3d; (b) the holes emerged form region growing are filled by applying the morphological close operations.

CHAPTER 4

4 Machine Learning and Data Analysis in Medical Data

Incisional hernia is one of the most common complications after abdominal surgery with an incidence rate of 11 to 20% post laparotomy. The impact in terms of Quality of Life (QoL) worsening and public health costs is relevant. Moreover, the recurrence of incisional hernia after surgical operation of incisional hernia repair has an incidence rate of 18 to 50% during the first post-operative year depending on some variables [38][39]. Particularly, many different factors can be considered as risk factors of incisional hernia recurrence such as the surgical technique performed, the use of a mesh, the anatomical site of the mesh positioning, the emergency regimen of the surgical operation, the presence of a dirty/contaminated surgical field, the oldness, the sex, the obesity, the diabetes mellitus type 2, the tobacco use, the malnutrition, the use of immunosuppressor, the chronic pulmonary disease, the ascites and the chronic anaemia [40].

Crohn's Disease (CD) is a life-long idiopathic, often debilitating, chronic inflammatory disease of the gut, which can potentially involve the entire gastrointestinal tract [1]. It arises from an interaction between genetic and environmental factors, and it is characterized by granulomatous autoimmune reaction of the bowel walls, from the mucosal layer to the serosa one, and frequent extraluminal and extra-intestinal features. The incidence peak is in adolescents and young adults, between 15 and 25 years old [41][42]. CD comprehends a variety of complex phenotypes in terms of disease location and behavior, characterized by different events within time, healing and relapses. This phenotypical heterogeneity depends on many factors: the age of appearance, time elapsed from symptoms to diagnosis, site, extent and behavior of disease, and other anamnestic features [43]. Because of these heterogeneities the last European Consensus by European Crohn's and Colitis Organization (ECCO) agreed on the lack of a single gold standard for diagnosis of CD: "A single gold standard for the diagnosis of CD is not available. The diagnosis is confirmed by clinical evaluation and a combination of endoscopic, histological, radiological, and/or biochemical investigations" [44]. The recent new therapies, both biological ones and immuno-modulators, are very effective in downgrading disease activity and in symptoms control, but their real effect on disease course is mostly unknown. Moreover different therapy strategies must be tailored on disease evolution during treatment, thus allowing to prolong remission, to improve life quality, to prevent or manage hospitalization and surgery, and finally to prevent disability [43]. Imaging plays a pivotal role on disease evaluation during life-time, allowing a non-invasive patients follow up. That's the reason why non ionizing

imaging technics must be preferred [43][45]. Enterography magnetic resonance imaging (E-MRI) was recently yielded as useful diagnostic tool that can afford radiation free diagnosis where other imaging examinations have failed, or when involved areas are not easily reachable by endoscopy [42][43][46]. It is considered the gold standard for CD diagnosis and is not able to provide a comprehensive assessment of extra-luminal features and extraintestinal manifestations of disease [43]. In the last decades, MRI represented a valid technique for evaluation of disease extension and activity, for patient follow-up or for a pre-operative assessment in patients with histologically proven CD [43]. Its role as first step examination in diagnosis for suspected but not confirmed CD was explored only recently, showing high accuracy (sensitivity 93%, specificity 90%) [47]. Up to date, many studies have been published, concerning the use of imaging evaluation in CD diagnosis and grading, although nowadays there is not yet an adequate consensus about its actual reliability in clinical practice [44][48][49]. Because of the widely heterogeneous clinical features of CD, the role of radiologists may be challenging [43][46][50][51]. Nevertheless, worldwide literature has reported typical E-MRI features, which are frequently associated with CD, and whose evaluation may help in diagnosis by imaging [43][50][51]. The possibility of early CD diagnosis has great clinical implications, considering the impact on public health that such a chronic and disabling disease can imply, due to the high economic costs necessary for patients management, and which may be potentially reduced by imaging early diagnosis, and pattern definition [43]. This is particularly important considering that time elapsed from symptoms to diagnosis, site, extent and behavior of disease deeply influence patient prognosis, outcome, therapy, and complications.

Automatic diagnosis of CD is a very challenging task and can have a great clinical implication in CD affected patients. In last years, a lot of approaches to classify MRI images into healthy or unhealthy classes have been proposed. Supervised classification techniques include support vector machine (SVM) [52] and k-nearest neighbors (k-NN) [53][54]. Unsupervised classification techniques [55][56] include self-organization feature map (SOFM) [52] and fuzzy c-means [54][57]. Generally, all these methods achieve good classification results.

Among supervised classification methods, SVMs are common classification methods based on machine learning [58][59][60]. Compared against other methods such as artificial neural networks, decision trees, and Bayesian networks, SVMs have significant advantages in high accuracy, elegant mathematical tractability, and direct geometric interpretation. Moreover, they do not need a large number of training samples to avoid over-fitting [61]. To the best of our knowledge, there are not SVM based applications for CD in literature. In [62], it has been applied a SVM method for data classification. The SVM modeling is a promising classification approach for predicting medication adherence in Heart Failure (HF) patients. This predictive model helps to stratify the patients so that evidence-based decisions can be made and patients can be managed appropriately. In [63], a novel hybrid system to classify a

given MR brain image as either normal or abnormal is proposed. The method employed digital wavelet transform to extract features and principal component analysis (PCA) to reduce the feature space. Afterwards, a KSVM with Radial basis function (RBF) kernel, using particle swarm optimization (PSO) to optimize the parameters C and σ is adopted. Five-fold cross-validation was utilized to avoid over-fitting. Wavelet transform is an effective tool for feature extraction from MR brain images, because it allows images analysis at various levels of resolution due to its multi-resolution analytic property. This technique requires large storage and it is computationally expensive [64]. PCA is appealing since it effectively reduces data dimensionality and computational cost [65]. In this paper a supervised Crohn's disease patient's classification technique based on Kernel Support Vector Machine (KSVM) algorithm using a Stratified K-Fold Cross Validation strategy is presented. The proposed technique aims to apply a KSVM technique on E-MRI qualitative extracted features to simplify classification task complexity and preserve technique accuracy and quality when compared against manual methods [47].

4.1 Use of the KSVM-based system for the definition, validation and identification of the incisional hernia recurrence risk factors

The proposed intelligent data analysis and classification technique are based on machine learning algorithms used to classify the proprietary University of Palermo Policlinico Hospital (UPPH) Dataset. The dataset is composed of 154 patients, each of one codified by 34 qualitative features and classifier into Positive for recurrence risk or Negative for recurrence risk by expert surgeons. The purpose of the study is reached developing an objective tool for incisional hernia affected patients classification, using a Kernel Support Vector Machine (KSVM) based technique. The proposed technique uses 103 vectors composed of 34 qualitative features used for KSVM training to identify incisional hernia recurrence; 51 vectors com- posed of 34 qualitative features used to validate the proposed technique. In order to make reliable the classifier, generalized to other sets of independent data and to limit problems as over-fitting, without increasing the precision of final standings, the cross- validation is integrated into the proposed technique. In Stratified K-fold cross-validation, the data set is divided into k equal subsets, and the holdout method is repeated k times. Each time, one of the k subsets is used as the validation set, and the other K-1 subsets are put together to form a training set. Then the average error across all K experiments is computed. In this work K has been empirically determined as 5 through the trial-and-error method.

4.2 A Kernel Support Vector Machine Based Technique for Crohn's Disease Classification in Human Patients

The purpose of this study is to develop an automatic tool for CD affected patients classification based on a kernel support vector machine (KSVM) technique. The

description of algorithms and mathematical formalisms used are briefly described.

4.2.1 Features Extraction

Parameters extraction has been elaborated on the basis of the typical E-MRI features of CD affected patients [43][46][50][51], as shown in Figure 4.2.1 and Table 4.2.1.

Table 4.2.1 MR Enterorgraphy extracted features.

Sequence types	Features	Parameters
T2 SPAIR/post contrast T1 e-Thrive/BFFE	Termina Ileum thickening	0: None or less than 3 mm thickening 1. Thickening greater than 3 mm
	Length	The length, in cm, of the affected gastrointestinal tract/tracts
	Single lesion/Skip lesions	1: single tract involved: 2: multiple tracts involved
	Fat wrapping	0: Normal mesenteric adipose tissue 1: mesenteric hyperplasia
	Sinus	0: No 1: Yes
	Fistulas	0: No 1: Yes
	Surgery	0: No 1: Yes
	Mucosal layer	0: no mucosal involvement 1: Edema and post-contrast enhancement 2: Inflammatory changes and pseudo-polyps
	Free fluid	0: No 1: Yes
	Pattern	0: no disease features 1: active inflammatory subtype 2: fibrostenotic subtype 3: fistulizing subtype
	Activity	0: no activity 1: mild activity 2: moderate activity 3: severe activity
	Complications	0: No 1: Yes
	Intestinal obstruction	0: No 1: Yes
Post-contrast T1 e-Thrive	Lymph nodes	0: No 1: Yes
	Breathing/ Peristalsis artifacts	0: No 1: Yes
	Post contrast T1 imaging	0: no wall enhancement 1: layered enhancement 2: transmural enhancement
T2 SPAIR	Pseudo-polyps	0: No 1: Yes
	Lumen	0: No changes in lumen caliber 1: Stenosis/Sub-stenosis
	T2W imaging	0: no mural edema 1: mild mural edema 2: severe hyper intensity due to noticeable edema
HASTE thick slab	Bowel cleaning protocol	0: no sufficient preparation 1: adequate bowel loops cleaning and distention
	Bowel distention protocol	0: no sufficient bowel distention: only stomach or duodeno-jejuneal loop 1: PEG has reached the ileocecal junction
DWI	Water diffusion restriction (DWI)	0: free and physiological diffusion, no hyper intensity on DWI 1: mild hyper-intensity 2: Severe hyper intensity

4.2.2 Kernel SVM

The support vector machines (SVM) can be thought as an alternative technique for polynomial classifiers learning, as opposed to classical training methods for neural networks. SVMs can represent complex non-linear functions. The characteristic parameters of the network are obtained by the solution of a quadratic convex programming problem with equality constraints or box type (in which the value of the parameter must be maintained within a range), providing a single global minimum [66]. One of the most popular and reliable SVMs are the kernel SVMs. Kernel SVMs have the following advantages [68]:

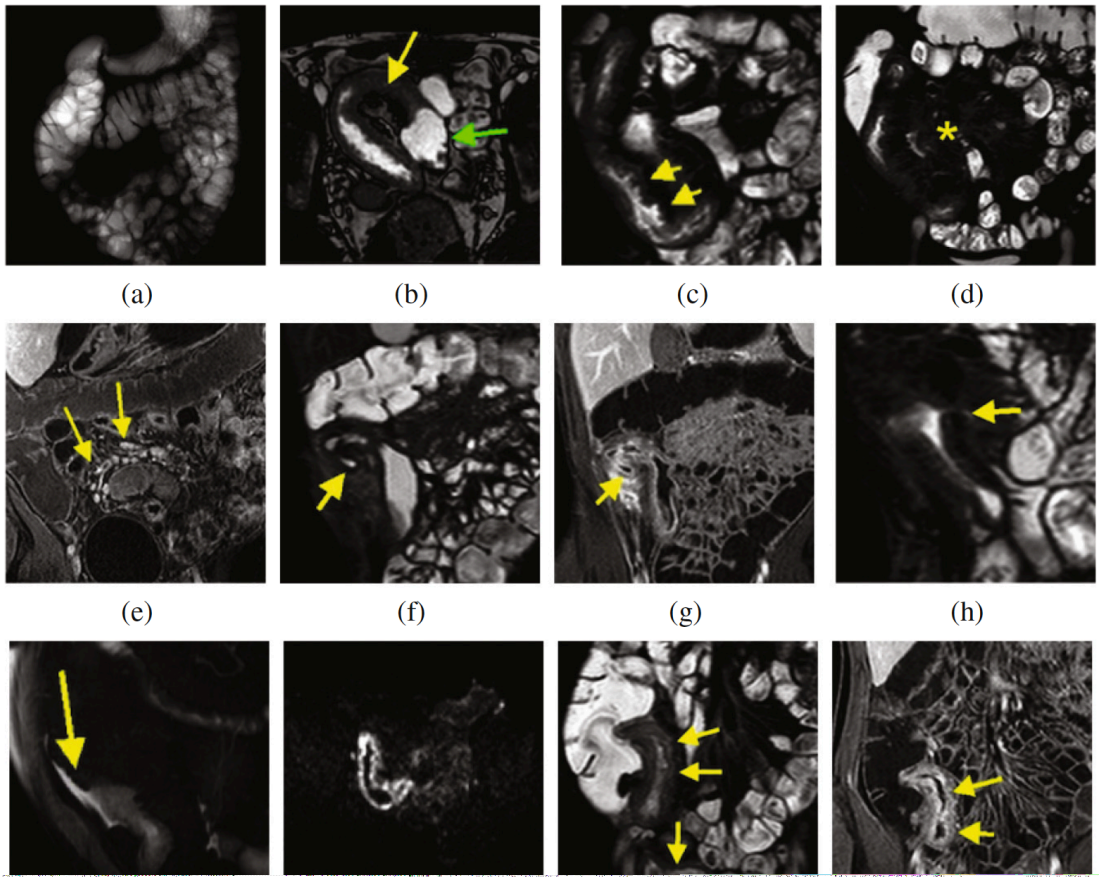


Figure 4.2.1 E-MRI CD Features: (a) HASTE Thick slab: Bowel cleansing and distension;

(b) Balanced - TrueFISP: Terminal Ileum thickening; length; lumen caliber: substenosis; (c) T2 Single-shot FSE fat-sat: Mucosal layer: pseudo-polyps; (d) T2 Singleshot FSE fat-sat: Fat Wrapping; (e) Post contrast T1 3D spoiled GE: Lymph-nodes; (f) T2 Single-shot FSE fat-sat: Sinus; (g) Post contrast T1 3D spoiled GE: Sinus; (h) T2 Single-shot FSE fat-sat: Fistula; (i) T2 Single-shot FSE fat-sat: Free fluid; (j) DWI: Water diffusion restriction. Hyper-intensity; (k) T2 Single-shot FSE fat-sat: T2 imaging: Hyper-intensity due to severe edema and Terminal Ileum Thickening; (l) Post contrast T1 imaging 3D spoiled GE: Layered pattern contrast

enhancement.

- Work very well in practice and have been remarkably successful in such diverse fields as natural language categorization, bioinformatics and computer vision;
- Have few tunable parameters;
- Training often involves convex quadratic optimization.

Four common kernels [67] have been used in this work and they are listed in Table 4.2.2. For each kernel, there should be at least one adjusting parameter to make the kernel flexible and tailor itself to practical data. Table 4.2.2 depicts also the used parameters in each case.
Our task is a two class classification task, labelled as Positive or Negative.

Table 4.2.2 Four common kernels used in the Support Vector Machine

SVM Kernel	Formula	Parameter
Linear	$K(x,y) = x^T y, x, y \in \mathbb{R}^d$	*Linear kernel, meaning dot product*
Polynomial	$K(x,y) = \left(x^T y + r\right)^n, x, y \in \mathbb{R}^d, r > 0$	*Polynomial kernel (order* $n = 3$*)*
Quadratic	$K(x,y) = \left(x^T y + r\right)^2, x, y \in \mathbb{R}^d, r > 0$	*Quadratic kernel*
Radial basis function (RBF)	$e^{\frac{\|x-y\|^2}{2\sigma^2}}, x, y \in \mathbb{R}^d, \sigma > 0$	Gaussian Radial Basis Function kernel (*scaling factor* $1/2\sigma^2 = 1$)

4.2.3 The Proposed Technique

The goal of the proposed system is the supervised Crohn's disease affected patient classification using a KSVM algorithm. Accordingly to the expert radiologists, each E-MRI sequence is codified by a vector composed of 22 qualitative features as depicted in Table 4.2.2. KSVM classification performs two class classification task: accordingly to the related histological specimen result each pattern is classified as Positive or Negative. As pointed out before, our dataset is composed of 300 vectors of 22 qualitative features. The whole dataset is divided into two bins: 80% (240 vectors) of the sampled values are used for the training session, and the remaining 20% (60 vectors) values are used for the validation session:

- 240 vectors composed of 22 qualitative features for KSVM training;
- 60 vectors composed of 22 qualitative features for KSVM validaion.

A cross-validation strategy has been also integrated into the proposed technique to enhance classifier reliability and generalization capabilities, and limit over-fitting issues. The proposed training and validation workflow is depicted in Figure 4.2.2.

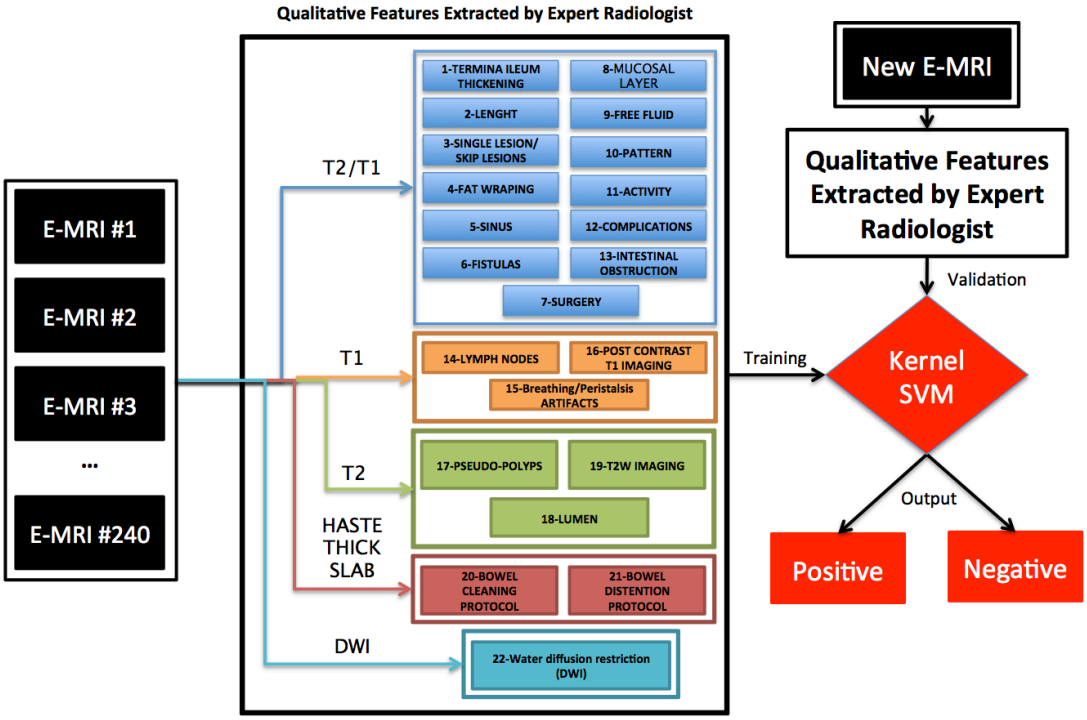


Figure 4.2.2 The proposed training and validation workflow: each E-MRI sequence is codified by a vector composed of 22 qualitative features. Accordingly to the related histological specimen result, KSVM training vectors have been labelled. Afterwards, a new E-MRI vector is classified by KSVM.

CHAPTER 5

5 Image Segmentation in PET

In radiation therapy planning, Magnetic Resonance Imaging (MRI) provides high quality detailed images and excellent soft-tissue contrast, while Computerized Tomography (CT) images provides attenuation maps. In this context, Positron Emission Tomography (PET) is a non-invasive imaging technique that has the advantage, over morphological imaging techniques, of providing functional and metabolic information about the patient's condition. Inclusion of PET images in radiotherapy protocols is desirable because it provides crucial information to accurately target the oncological lesion and to escalate the radiation dose without increasing normal tissue injury [69]. For this reason, PET may be used for improving the Planning Treatment Volume (PTV) [70]. In addition, PET quantitative assessment in oncological patients, conveys functional information which is predictive of treatment response [71] and faster changing than the morphological response [72]. As such, PET can greatly improve the clinical cancer treatment decision making [3]. In particular, the maximum Standardized Uptake Value (SUVmax) is the most widely used quantification parameter giving a punctual measure of cellular metabolism [73]. To provide further information about the cancer, additional quantitative parameters have been introduced, such as biological tumor volume (BTV) and tumor lesion glycolysis (TLG) [74][75]. TLG is obtained as the product of SUV with BTV and provides both volumetric and metabolic information. Therefore, BTV delineation is one of the most important steps towards precise radiation therapy treatment planning and quantification accuracy [76].

In general, PET segmentation methods can be categorized as manual, semi-automatic and automatic. Manual contouring is still widely adopted in clinical environments because it is easily applicable, and because a large number of operators feel that no algorithm would segment "how they would". However, accuracy of manual BTVs may be affected by blurred contours due to the image resolution and also depends upon both the operator's expertise and clinical specialization. Additionally, manual segmentation is time consuming. Indeed, the level of performance imposed by daily clinical routines, makes properties such as repeatability (i.e. the result should be operator independent) and real-time processing not only desirable, but mandatory. Thus, automated or semi-automated methods are highly desirable in order to reduce inter-operator variation and time. Unfortunately, BTV contouring is strongly depend on the segmentation algorithm [77] and suffers from the typically low resolution of PET images [76][36]. As a result, the choice of a standard BTV segmentation algorithm remains a challenging, unresolved issue.

A large number of PET segmentation methods can be found in the literature [78][79]. Thresholding and region growing (RG) methods are among the most popular, principally because they are easily implemented.

These approaches however, show a drop in performance on low-contrast heterogeneous cancer regions [11] and segmentation of small or non-spherical tumors is often unsatisfactory [80]. The adaptive RG algorithm in [81] repeatedly applies a confidence connected RG algorithm with an increasing relaxation factor automatically identified. This algorithm is robust to parameter settings and region of interest selection in the case of a relatively homogeneous background. Additionally, the approach seems to be free from dependency on scanner type, imaging protocol, or tumor shape.

Affinity propagation [82] investigates multi-focal radiotracer uptake patterns. Unfortunately, its development is still limited to animal studies.

Stochastic models (e.g. the Gaussian mixture model), founded on the statistical analysis of the intensity distribution, can be considered optimal for noisy images. Nevertheless, proper noise models must be determined a priori [83].

Learning methods [84], such as artificial neural networks, and support vector machines are efficient, although the training of such algorithms typically requires large and diversified datasets.

Due to the fuzzy nature of the lesion contours in PET studies, the Fuzzy C-Means (FCM) [85] approach has become very popular for delineation. However, while large and simple-shaped targets are accurately recovered, complex-shaped lesions remain a challenge [86].

Variational approaches based on gradient differences between healthy and cancer tissues are designed for maximum mathematical efficiency, but are sensitive to noise and subject to numerical fluctuation [27]. Active contours (AC) belong to this family of algorithms. An initial contour, drawn around the object of interest is evolved toward the object edges. The evolution is mathematically driven by the gradient of what is termed as the “energy function”, a scalar function mathematically defined in such a way that its minimum value is attained by the best possible segmentation. Some examples are available in literature: Li et al [87] used RG as a pre-processing step to strategically initialize the active contour. Unfortunately, the result tended to overestimate the tumor volume [81]. Similarly, ACs have been used in combination with anisotropic diffusion filtering and multi-resolution contourlet transforms [88]. However, the outcome of this approach heavily depends on several user-defined parameters. In order to improve accuracy and robustness, Fuzzy C-means (FCM) clustering and textural information have been used to improve AC robustness [89]. Nevertheless, this strategy suffers in the presence of nearby high physiologic uptake, and its result depends on the initial cropping area.

Graph-based strategies yield efficient segmentation, the underlying rationale being to locate different tissues by using foreground and background seeds [90]. However, when seed identification is automated, high uptake physiological structures (e.g.

brain, heart, bladder, and kidneys) are often mistakenly identified as initial target seeds, resulting in misleading guidance to the segmentation algorithm [91].

Alternative studies exist which tackle the challenge of automatically discriminating between normal and pathological tissues in PET. Unfortunately, a full comparison of such algorithms is not possible as the relative studies often concern different body districts and specific types of abnormality. Studies on the discrimination of pathological structures in whole-body PET have been conducted as well, and some preliminary results have shown that different anatomical areas pose different challenges [92][93]. In addition, numerous PET-based radiometric studies have been proposed, and the results of the relative analysis are highly dependent on the method used to derive the BTV [94].

The combination of PET and CT or MRI in a single scanner is desirable, as it offers the advantage of generating complementary information. However, a one-to-one relationship between anatomical and functional images is not always appropriate [91]. Co-segmentation methods, which assume no significant anatomical and functional changes between the images acquired by different modalities, generally yield a single target volume where the information of each modality is not optimally exploited, or at worse, lost.

When PET/CT or PET/MRI imaging is employed, BTV delineation should be obtained, avoiding the use of the anatomical information provided by CT and MRI measurements, or alternatively, the latter should be used very carefully [95]. For example, Yezzi et al. [96]proposed active contours to simultaneously segment and register features from multimodal images.

Despite this flourishing population of segmentation algorithms, most operators still feel more comfortable with manual segmentation. The main reason is because contouring algorithms are likely to include portions of the tumor or surrounding tissue that the physician would had discarded, or they may overlook portions of tissue the physician would otherwise include. In practice, an expert human would have segmented differently by virtue of his or her own professional training and experience.

The main subject of the present work is an operator independent segmentation strategy to be used in radiotherapy (RTP).

Such a strategy is implemented in the form of an algorithm with the following key features:

1) preliminary pre-processing which converts PET DICOM data (Digital Imaging and Communications in Medicine) [97] into SUV images in order to incorporate important metabolic information;

2) minimal human intervention, limited to a rough contour to highlight an initial ROI on just a single PET slice, yielding considerable time reduction; and

3) two fully automated stages of segmentation. The first segmentation stage consists of the preliminary identification of a rough volume of interest from which the optimal starting mask is extracted. Such a mask consists of a user-independent

contour of the tumor boundary. It is individuated on the slice containing the maximum value of SUV for the whole lesion. As such it may actually lie on a PET slice which differs from the user drawn ROI. The second stage, which is initiated from the optimal mask, yields a high detail segmentation. This latter stage, in particular, follows a slice-to-slice walking strategy and implements an innovative algorithm obtained by coupling a local active contour (LAC) with tissue classification from the classification method [98][99].

The employed datasets focus on the uptake of 18F-fluoro-2-deoxy-d-glucose (FDG) and 11C-labeled Methionine (MET) radio-tracers. On the one hand, FDG (an analogue of glucose) accounts for approximately 90% of all PET imaging procedures. On the other hand, MET-PET has being recognized as an efficient means of highlighting tumor cell invasion in brain metastases, and according to some studies [100][101] may even outperform CT or MRI

5.1 A smart and operator independent system to delineate tumours in Positron Emission Tomography scans

In order to overcome the general limitations of the above mentioned studies, we combined and adapted to PET imaging several existing technologies. Doing so, we created a new smart system (Figure 5.1.1) whose main purpose is to deploy a segmentation strategy to be used in radio-therapy treatment planning, and outcome evaluation. The proposed system is applicable to various types of lesion, different anatomical districts and can perform segmentation on PET studies based on different radiotracers. Our approach reduces at minimum the intra-observer and inter-observer dependencies with respect to manual delineation, without introducing any user-dependent parameter. The AC segmentation algorithm at the core of the system is applied using a slice-by-slice approach, starting from an initial, automatically individuated slice. Process is terminated when a suitable stopping condition is met. Consequently, the system presents a high level of automation. In the following we provide a high level description of our approach.

Briefly, the region containing the cancer must be localized by the operator to avoid healthy tissues having high radio-tracer uptake or critical conditions, i.e. when the lesion is located near the heart. The PET Digital Imaging and Communications in Medicine (DICOM) dataset is then converted into SUV images to normalize the voxel activity, taking into account functional aspects of the disease. Based on this minimal input, where the user is required to highlight a region on just a single PET slice, the algorithm performs all subsequent operations automatically. The first step consists of a pre-segmentation that computes a user independent region of interest (ROI). The algorithm automatically finds an optimal starting mask, which may lay on a slice different from the one initially highlighted by the operator. Once the independent ROI has been obtained, the relative initial mask is feed to the next step

of the system, where the segmentation is performed using a Local region-based Active Contour (LAC) segmentation algorithm, appropriately modified to support metabolic images as explained in the following sections. The obtained LAC segmentation is propagated to the adjacent slices using a slice-by-slice marching approach. Propagation is performed in parallel both upward and downward within the SUV volume to obtain the BTV, until a suitable stopping condition is met. In the following, a detailed discussion of the various steps is presented.

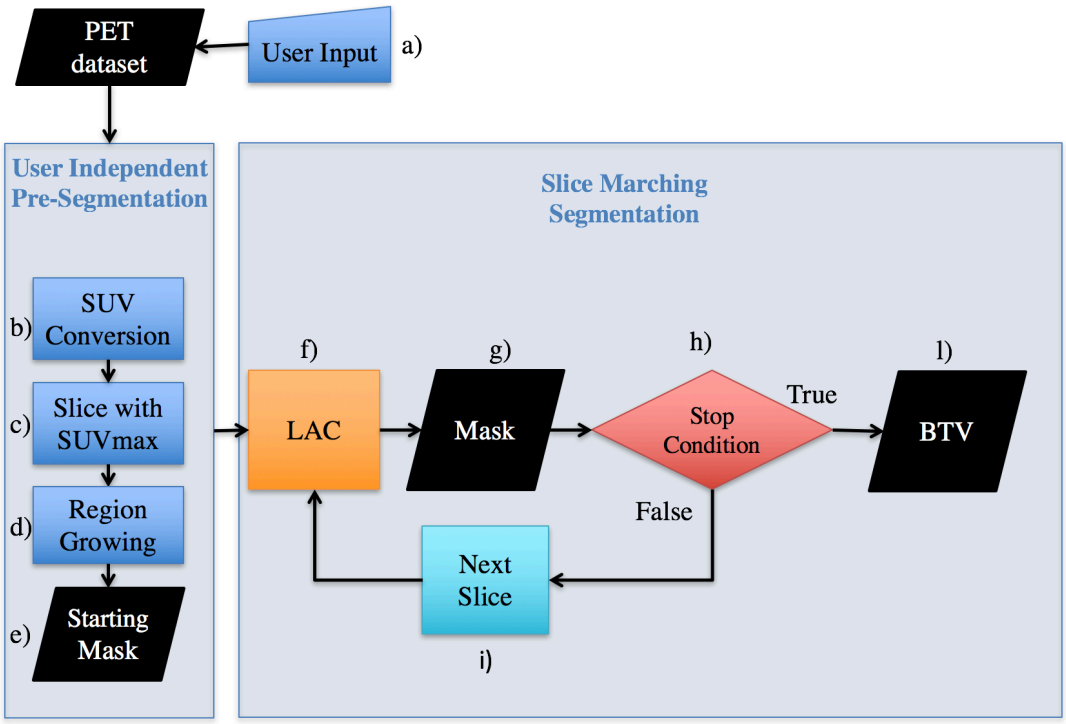


Figure 5.1.1 The proposed segmentation system. a) to avoid any false positive, the region containing the lesion is localized by the operator; b) PET dataset is converted into SUV images to incorporate functional information; c) The PET slice containing maximum Standardized Uptake Value is found d) RG is used to identify the user independent ROI on the slice; e) initial ROI is sent to the next logical block of the system; f) segmentation is performed using the LAC algorithm; g) segmentation mask is propagated to the adjacent slices (slice-by-slice marching approach); h) the stop condition is evaluated (see following sections); i) segmentation on the next slices is performed until stop condition is false; l) an operator independent BTV is finally obtained.

5.1.1 User Independent Pre-Segmentation Step

We defined the proposed system as semi-automatic because in total body FDG-PET examinations, an initial user input is still required to roughly identify the area containing the cancer. The motivation is to avoid false positives when cancer is

situated near anatomical structures such as brain, heart, bladder, kidneys, and ureters where high FDG uptake is to be considered normal. However, user intervention is minimal and limited at drawing a rough contour around the cancer area, on just one PET slice (Figure 5.1.1-a). No further intervention is needed. The input technique (e.g. use of touch screen, stylus, etc.) depend on user's preference. In the present study input was performed by mouse.

Pre-processing the uptake data in PET examinations is mandatory for inter-patient and follow-up comparisons. Among PET metrics, SUV is the most widely used quantification parameter giving a measure of cellular metabolism. For this reason, we integrated SUV information in our system. The PET dataset is converted into weighted SUV unit (g/ml) [102] images (Figure 5.1.1-b), as suggested in [86]. SUV is calculated as the ratio of tissue radioactivity concentration (RC) in kBq/ml and injected dose (ID) in MBq at the time of injection divided by body weight ($\mathrm{M_p}$) in kilograms, thereby taking into account essential functional aspects of the patient:

$$SUV_i = \frac{RC}{ID} * M_p$$

RC is calculated as the ratio between the image intensity and the image scale factor. ID is the product between actual activity and dose calibration factor [74][90].

Based on the region highlighted by the user (ROI0), the first logical block of our algorithm consists of an automatic pre-segmentation step which computes a user independent ROI. This goal is achieved by first identifying a volume of interest (VOI) which is formed by propagating ROI0 to adjacent slices.

Every time a new slice (index j) is considered, ROI0 is propagated to obtain ROIj, and the maximum SUV within ROIj is detected ($\mathrm{SUV^j_{max}}$). If this value exceeds 42% [95] of $\mathrm{SUV^0_{max}}$ found within the initial slice, then the new slice is added to the volume. A new ROIj is then computed in the slice j using the region growing (RG) method [96] to follow the evolution of the lesion along the SUV volume. In particular, the $\mathrm{SUV^j_{max}}$ voxel is used as target seed. Otherwise, the growing is stopped. In this way, a cylindroid (i.e. a cylinder with a non-circular base) is obtained.

The next pre-segmentation step is to locate the voxel of maximum SUV within this volume of interest, $\mathrm{SUV^{VOI}_{max}}$ (Figure 5.1.1-c). Note that this maximum value may reside on a different slice than the one selected by the user, but it will belong to the same anatomical anomaly.

The rough segmentation of the newly identified slice using the RG algorithm (Figure 5.1.1-d) represents the output of the pre-segmentation step (Figure 5.1.1-e) which will feed the next logical block of the system (Figure 5.1.1-f).

Note that in this pre-segmentation process, the region highlighted by the user does not need to be carefully drawn. A rough inclusion of the anomalous area is sufficient. Further, manual drawing does not need to be repeated on any other slice. As a result,

the pre-segmentation process generates an output that is independent to local variations in the initial user input, and for that, thereby making the result extremely repeatable. Furthermore, the RG algorithm is only used to obtain a rough estimate of the contour encircling the highest radio-tracer uptake area within the slice containing the SUV_{max}^{VOI}. This contour is then feed into the next section of the system, the "Slice Marching Segmentation" as shown in Figure 5.1.1, where the delineation is further refined through a more sophisticated segmentation algorithm (Figure 5.1.1f), accompanied by an automated data-driven stopping condition (Figure 5.1.1h).
Consequently, we can expect the final BTV (Figure 5.1.1) to contain the cancer within its limits and the segmentation to account for extremely irregular lesion shapes. Three examples of ROI bounding regions delineated on PET images by the RG method are shown in Figure 5.1.2(a–c).

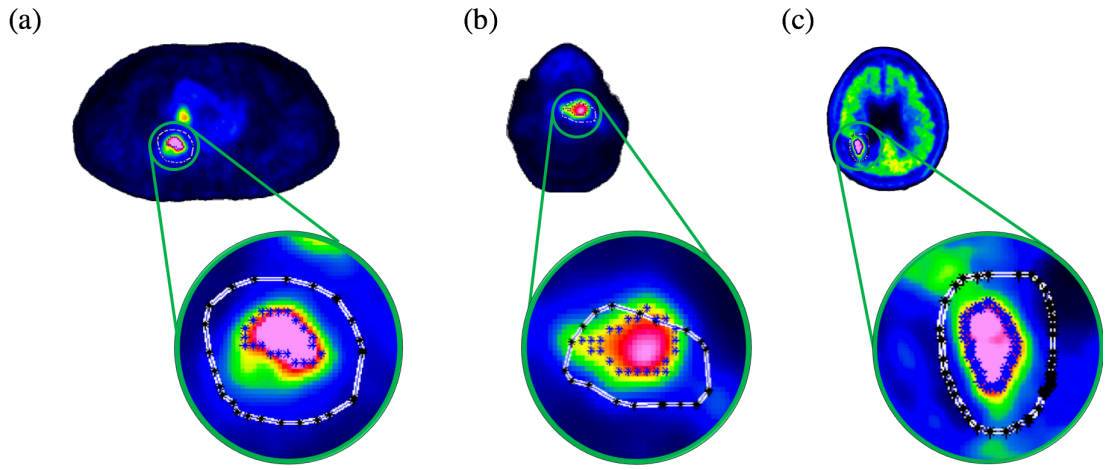


Figure 5.1.2 Operator independent ROI generation on PET images. Figures a, b, and c show the pre-segmentation output (Figure 5.1.1e) on three different body districts: lung (a), brain (b) and head and neck (c), respectively. The white contour with black asterisks represent the input provided by the user, while the blue asterisks contour shows the result of the RG algorithm after the slice containing the maximum SUV is identified. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.1.2 Slice Marching Segmentation Step

The LAC algorithm blends benefits of purely local edge based active contours and fully global region based active contours. At each point along a prominent intensity edge of an object, nearby points inside and outside the object will be modelled well by the mean intensities within the local neighbourhoods on either side of the edge. This is the motivation behind the model proposed by Lankton et al [83]; where the

contour energy to be minimized is defined as:

$$E = \oint_C \left(\int_{R_{in}} \chi_l(x,s)(\mathrm{I}(x) - u_l(s))^2 dx + \int_{R_{out}} \chi_l(x,s)(\mathrm{I}(x) - v_l(s))^2 dx \right) ds$$

where Rin and Rout represent the regions inside and outside the curve C; where *s* represents the arc-length parameter of *C*, where χ represents the characteristic function of the ball of radius l (local neighbourhood) centred around a given curve point *C*(*s*), where I represents the intensity function of the image to be segmented, and where $u_l(s)$ and $v_l(s)$ denote the local mean image intensities within the portions of the local neighbourhood $\chi_l(x, s)$ inside and outside the curve respectively (within Rin and Rout). These neighbourhoods are defined by the function χ, the radius parameter l, and the position of the curve C. Note that the function $\chi_l(x, s)$ evaluates to 1 in a local neighbourhood around each contour point C(s) and 0 elsewhere, thereby localizing the processing of image information. The shape of the contour C then divides each such local region into interior local points and exterior local points in accordance with the contour's role to segment the domain of I. The resulting flow is more robust to initial curve placement and image noise like region-based flows described on [103]; and yet it also capable of capturing significant local structure and partitioning the image without making strong global assumptions about its makeup. To be successful, the present segmentation technique relies on the assumption that nearby points inside and outside the true edge of an object, will be modelled well by the mean intensities within these localized regions. As a result, the energy (2) is more global in nature than edge-based flows but with a "tunable" degree of locality defined by the neighbourhood radius l.

Actually, the LAC method for 3D MR datasets was applied by Lankton et al [83] via independent segmentation of the 2D slices. A more powerful and coherent segmentation procedure could be performed on all slices simultaneously by evolving a single surface within the corresponding three-dimensional space. While we are currently investigating such 3D shape evolution for future evaluation and publication, the present work moves a step toward 3D data segmentation improving upon Lankton et al by introducing a system to automatically govern the slice-by-slice strategy. Further improvements consist of introducing the SUV measurements (functional information), an optimal identification of the starting slice, the production of an initial mask for LAC segmentation, and a fully automatic stopping condition. The latter, in particular is a key feature in order to achieve automation.

We now describe these new improvements integrated into the LAC algorithm.

In step f) (see Figure 5.1.1), an initial mask is obtained from user independent smart pre-segmentation (seen previous sections). To incorporate SUV in the LAC algorithm, the energy (2) for the PET image segmentation approach is adapted as:

$$E = \oint_C \left(\int_{R_{in}} \chi_l(x,s)(\mathrm{SUV}(x) - u_l(s))^2 dx + \int_{R_{out}} \chi_l(x,s)(\mathrm{SUV}(x) - v_l(s))^2 dx \right) ds$$

where SUV represents the intensity function of the standardized uptake value to be segmented, and where $u_l(s)$ and $v_l(s)$ denote the local mean SUV intensities within the portions of the local neighbourhood $\chi_l(x,s)$ inside and outside the curve respectively (within Rin and Rout). These neighbourhoods are defined by the function χ, the radius parameter l (in our study l=3 has been determined to provide the best performance using trial and error strategy, see appendix), and the position of the curve *C*. Note that the function $\chi_l(x,s)$ evaluates to 1 in a local neighbourhood around each contour point *C*(*s*) and 0 elsewhere, thereby localizing the processing of SUV image information. The shape of the contour *C* then divides each such local region into interior local points and exterior local points in accordance with the contour's segmentation of the SUV. The local means are specified in terms $S_{I_l}(s)$, $S_{E_l}(s)$, $A_{I_l}(s)$, and $A_{E_l}(s)$ which represent the local sums of SUV intensities and the areas of their respective portions of the local neighbourhood $\chi_l(x,s)$ inside and outside the curve (within Rin and Rout). More precisely, the local interior region may be expressed as $R_{in} \cap \chi_l(x,s)$ and local exterior region as $R_{out} \cap \chi_l(x,s)$.

$$u_l(s) = \frac{S_{I_l}(s)}{A_{I_l}(s)} \quad , \quad v_l(s) = \frac{S_{E_l}(s)}{A_{E_l}(s)}$$

$$S_{I_l}(s) = \int_{R_{in}} \chi_l(x,s) SUV(x)\, dx \quad , \quad S_{E_l}(s) = \int_{R_{out}} \chi_l(x,s) SUV(x)\, dx$$

$$A_{I_l}(p) = \int_{R_{in}} \chi_l(x,s)\, dx \quad , \quad A_{E_l}(p) = \int_{R_{out}} \chi_l(x,s)\, dx$$

$$\chi_l(x,s) = \begin{cases} 1 & when \;\; x \in l - Ball(C(s)); \\ 0 & otherwise; \end{cases}$$

After the segmentation step f) is achieved on the slice containing the $\mathrm{SUV}_{\mathrm{max}}^{\mathrm{VOI}}$, the resulting segmented mask is used to initiate parallel segmentations on the

neighbouring slice above and below. Subsequently, for all the other slices in both directions, we similarly use the segmentation results of the previous slices as the initial mask inputs as shown in step g) Figure 5.1.1. The LAC method is inherently capable of locally widening or tightening where necessary when the contour is propagated from slice to slice. Since, this behaviour is driven by the image properties rather than by an inherent knowledge of whether the cancer is present, a stopping criterion is necessary to prevent the LAC algorithm from misbehaving or even diverging when it reaches a slice where the cancer is absent (i.e. when there is nothing to be segmented).

Therefore, we devised a fully automatic stopping condition as shown in step h) Figure 5.1.1. For the slice under consideration, at each point on the cancer edge, nearby points inside and outside the cancer must have a different local mean SUV. If the cancer is present, a positive difference between background and foreground intensity must occur, and consequently the algorithm can safely proceed with the next neighbouring slice (Figure 5.1.1-i). When the system encounters a slice where the local mean v_l (s) on Rout is greater or equal to the local mean SUV u_l (s) on Rin , which is the opposite of what is expected, the slice is classified as cancer-free and the slice-to-slice segmentation propagation is halted (Figure 5.1.1-l) in that direction. The approach can be mathematically summarized as:

$$\text{Stop Condition} = \begin{cases} \text{segmentation process is stopped} & when \;\; v_l(s) \geq u_l(s) \; is \; true; \\ \text{segmentation process is continues} & otherwise; \end{cases}$$

In this way, slice after slice, the BTV is obtained and the segmentation process is automatically stopped, thereby avoiding the need for any user intervention. In the following section, the criteria to judge the quality of the segmentation results is outlined.

5.2 Normal and abnormal tissue classification in positron emission tomography oncological studies

The aim of this study is to develop an objective tool for tissue classification using a KNN based technique. A dataset of 80 oncological patients is collected, and for every patient, lesion (unhealthy tissue) and back- ground (healthy tissue around the lesion) are contoured on PET images using fixed 2D ROI of 5 × 6 voxel size. The ROI size has been determined by trial-and-error methodology: the 30 voxel size provided the best result among those evaluated in tissue classification; the size ranged from 9 (ROI of 3 × 3 voxel size) to 49 (ROI of 7 × 7 voxel size) voxels. Then, a 30- dimensional vector is obtained by concatenation of the voxels within the ROI. Each vector is composed of 30 SUVs characterizing the area under investigation (lesion or background). Finally 160 vectors are used to train and validate the KNN.

5.2.1 Proposed Method

KNN is used as system to classify normal and abnormal regions in PET oncological studies for assisting physicians in radiation treatment planning. No parameters are requested from the operator. The main task is learning to classify tissues into two classes; the goal is to classify in which class a new tissue must be inserted. Starting from a semi-automatic method, 160 vectors characterizing the area under investigation are extracted from 80 oncological patients and used for training and validation purpose. To automatically label the ROI as abnormal or normal tissue, the nuclear medicine physician must place a fixed ROI, consisting of a red and a green section of 30 voxel size, along the lesion boundary so as to separate the target from the background region, as shown in Figure 5.2.1.

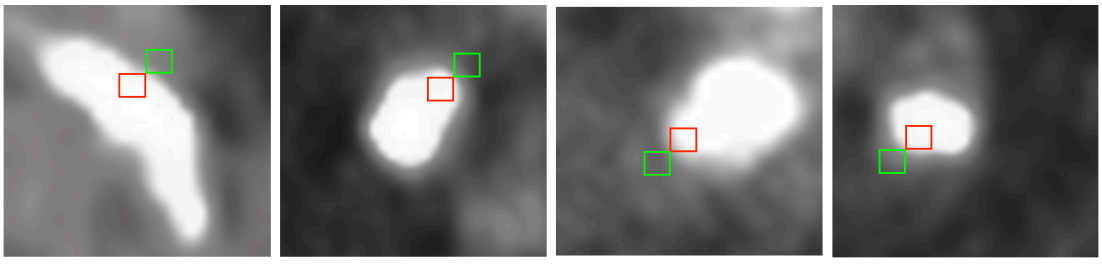


Figure 5.2.1 Fixed ROI, consisting of two rectangles, is placed along the lesion boundary so as to separate the target from the background.

The red section must be placed in the lesion, the green one in the background region. The physician can choose between two fixed ROIs (see the difference between fixed ROIs in Figure 5.2.1a, b versus Figure 5.2.1c, d): in this way, ROI can be correctly positioned and regions are correctly labeled. Hence the PET image is converted into a SUV image to obtain training vector composed of 30 SUVs characterizing the area under investigation. The ROI labels (abnormal or normal tissue) provided by the semi-automatic method are used as Ground-truth. Then the ROI set is divided into two bins: 80% (128 vectors) of the sampled values are used for the training session, and the remaining 20% (32 vectors) values are used for the validation and test purposes.

The K-Fold cross-validation has been integrated into the proposed technique to make reliable the classifier and to generalize the other sets of independent data and to limit problems such as over-fitting, with- out increasing the precision of final standings.

The error of the resulting estimate and the calculation time are reduced as K is reduced and vice versa [94]. In this work K has been empirically determined as 5 through the trial-and-error method (k range: 5–15, step size of 5). Finally, the optimal K value corre- sponding to the highest classification accuracy has been selected.

In one case, training vectors are labeled as normal or abnormal tissues using the semiautomatic method; in other cases, FCM and k-means are used for label- ling vectors in an unsupervised manner [54]. Briefly, the k-means algorithm is applied to partition 160 observations, where each observation is a 30-dimensional real vector, into two clusters in which every observation belongs to the cluster with the nearest mean. The fuzzy c-means is applied in the same way of the k-means algorithm.
When the training stage is complete, the validation step is started: 32 vectors are supplied to the input at KNN trained kernel for classifying tissue vectors as normal or abnormal. Finally, the KNN is ready and able to classify each new tissue vector in input using a single fixed ROI of 30 voxel size centered in the tissue of interest. The proposed system is shown in Figure 5.2.2.

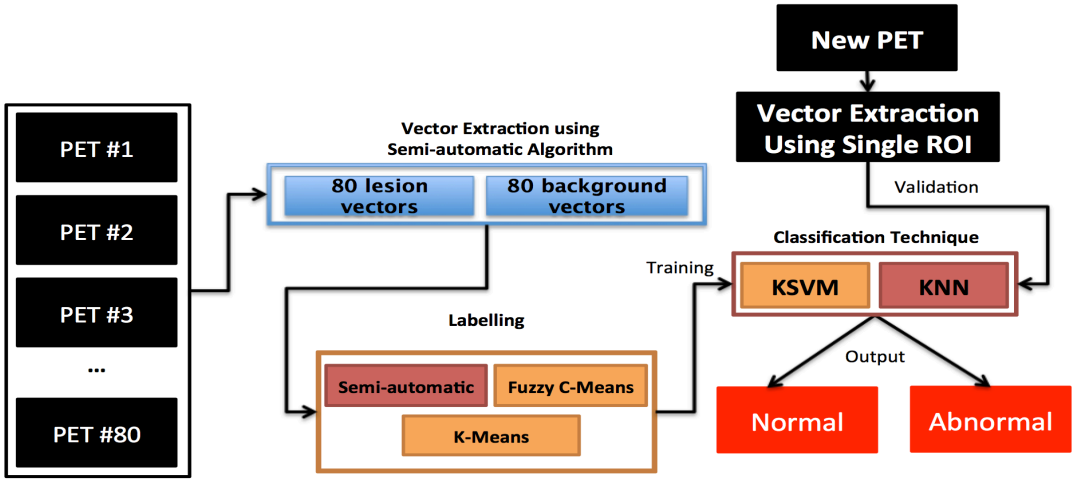


Figure 5.2.2 The proposed system: 160 vectors composed of 30 SUVs are labeled using a semiautomatic method as abnormal or normal. K-means and FCM are also used to label the vectors in an unsupervised manner. KNN and SVM are used to classify the input.

5.3 K-nearest neighbor driving active contours to delineate biological tumor volumes

To provide the reader a high level understanding of the proposed system, we can conceptually partition our algorithm (Figure 5.3.1) into four blocks. The first, "Sampling, Training and Tissues Classification" consists of the training and performance evaluation of the classifier. Since our goal is to combine the tissue labeling derived from KNN classification with the image information which locally drives the LAC segmentation, a preliminary training phase is required. This step provides the KNN with the capability of efficiently classifying a newly-encountered scenario (i.e. new tissue) into three tissue classes. However, this task needs to be performed only once. Once trained, the classifier is ready to be used in new clinical cases. Since this offline operation is carried out beforehand, we will discuss it first (subsection 2.1). Blocks 2, 3, and 4 followed in sequence. Block 2 "Initialization", is where the PET data are roughly prepared for processing and is the only block where

(minimal) user input is required. Block 3 "User Independent Pre-segmentation", implements a simple strategy to identify the optimal starting point for the next logical block of the system.

Finally, the fourth and last block "Slice Marching Segmentation", is where the true segmentation takes place, by the combined use of the KNN and LAC.
More detailed explanation of each block and its design is provided within the following subsections.

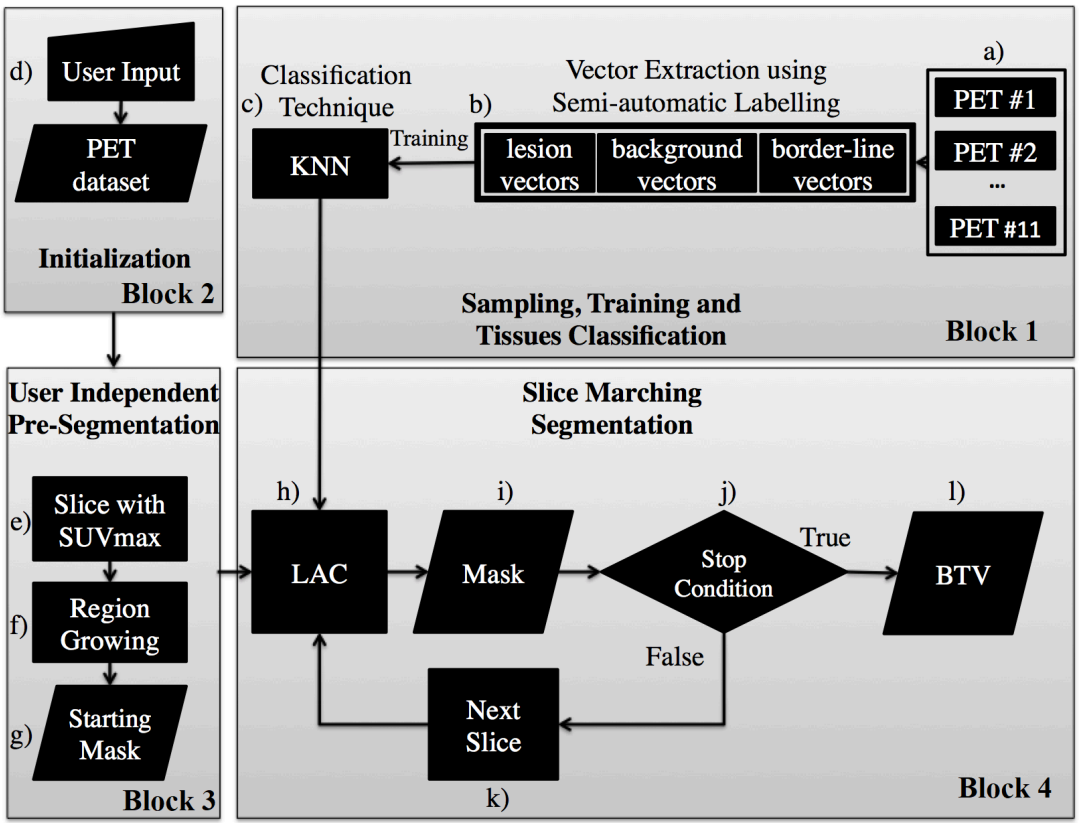


Figure 5.3.1 The proposed segmentation system. a) eleven different cases (8 patient lesions and 3 phantom spheres) are jointly processed by three expert operators; b) sampling operation is obtained for each tissue kind label; c) training and validation of the KNN classifier; d) the region containing the lesion is localized by the operator to avoid any false positive; e) global SUV_{max} is found after a pre-segmentation step; f) RG is used to identify the user independent mask; g) once the initial mask is obtained, it is sent to the next logical block of the system; h) segmentation is performed using the enhanced LAC segmentation algorithm through a new energy based on the KNN classification; i) segmentation mask is propagated to the adjacent slices using a slice-by-slice marching approach; j) the stop condition is evaluated (see subsequent sections); k) segmentation on the next slices is performed until stop condition is false; l) an operator independent BTV is finally obtained.

5.3.1 Block 1: Sampling, Training and Classification using KNN Algorithm

The idea behind the use of a classifier is to incorporate some of the physicians'

knowledge into the segmentation in the form of a trained software component. A classifier implementing just three classes, such as “lesion”, “background”, and “border-line” tissue constitutes a simple and suitable choice.
Proper training and validation are essential prerequisites to obtain reliable results. In order to generate the training input to the classifier, the clinical staff, consisting of the chief nuclear medicine physician, the chief radiotherapy physician and an expert radiotherapy physician (M.I., M.S., and G.R. authors, respectively), were required to place fixed-shape ROIs in a subset of 11 different cases (three phantom spheres and 8 patient lesions) from the initial PET dataset (Figure 5.3.1-a) with the purpose of identifying three different areas according to the classes being sought. In particular, the operators had to place a fixed ROI consisting of three different regions along the lesion/sphere boundary. Each region consisted of 7 by 7 voxels. The ROI size was empirically determined to provide the best performance on the present dataset with the specific choice of classifier. In this way, samples of grey levels from unhealthy tissue (red), healthy tissue (green), and the tissue immediately surrounding the lesion (yellow) were obtained as shown in Figure 5.3.2. Each sample was then reorganized as a 49 element vector. The sampling operation resulted in a total of 462 vectors, 154 for each type of tissue (Figure 5.3.1-b), considering fourteen samples for each patient\phantom study. Fixed-shape ROIs were located in the whole PET volume containing the target (lesion or sphere) for each study; they did not have to be placed in the same axial slice. To provide the classifier the capability of labelling the tissues in the most general case, the training included various body districts (two brain metastases, four HNC, and two lung cancers) and three phantom spheres with differing foreground/background radioactivity concentration ratios (3:1, 5:1, and 8:1) and differing diameters (28, 22, and 17 mm, respectively). Eighty percent (370 on 462 vectors) of the samples were used for training, while the remaining twenty percent (92 on 462 vectors) were used to check the classifier performance (Figure 5.3.1-c).

The rationale for the use of a KNN algorithm [93][20] is founded on the chosen sampling approach. We used N-by-N portions of the image taken from the slices as samples. The physicians’ expert segmentations provided the labeled training samples. In general, a sample residing within an obviously healthy region will exhibit low and almost constant uptake across the sample area.

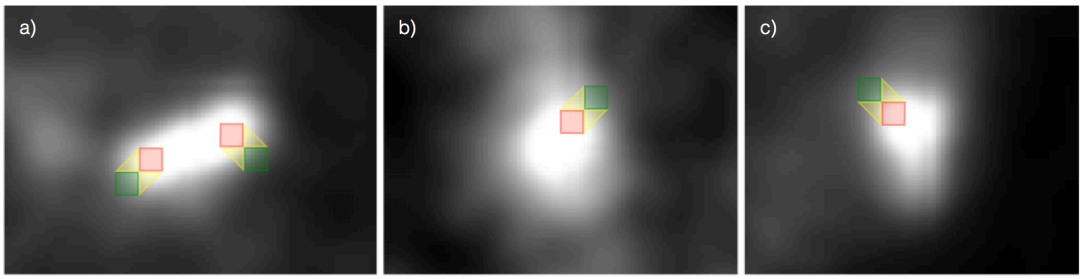


Figure 5.3.2 Fixed ROIs consisting of three regions (red, green, and yellow areas) are placed along the lesion boundary so as to separate the target (red) from the background (green). The yellow area indicates the tissue immediately surrounding the lesion. Three cancer regions are shown: HNC (a), brain (b), and lung (c).

The same will happen for the tumor, but with much higher average uptake (this is exploited by threshold method). On the one hand, these features strongly encourage KNN similarity when a new sample is provided for classification. The "border line" class, on the other hand, is the most interesting part. Such samples are found along the border drawn by the MD. They typically present fairly inhomogeneous SUV uptake values, and tumor and healthy tissue will ideally occupy opposite sides (although with a smeared transition). In other words, the three classes have strongly distinct features, and a KNN is expected to perform well in such a context.
The guidance provided by the classifier to the segmentation algorithm is implicitly founded on such a "border line" class and this is, in turn, what makes the difference between the pure (mathematical) LAC and our approach.
In practice, the KNN was selected because it provided the simplest and, given our sampling approach, the most advantageous alternative. Of course, other classifiers might be used as well and comparison with alternative algorithms will be discussed in future studies.

To strengthen the reliability of the results and to avoid common issues such as over-fitting, K-Fold cross-validation has been integrated in the KNN [93][94]. When the training stage is completed, the validation step is started: 92 vectors are input to the trained KNN kernel to verify the rate of success in classifying the new scenarios.

5.3.2 Block 2: Initialization

As a preliminary action, the PET data intensity values are converted into SUV units (g/ml). In the PET context, SUV is perhaps the most widely used quantitative measure of the cellular metabolism. SUV is calculated, as suggested in [86], by:

$$SUV_i = \frac{RC}{ID} * M_p$$

where RC denotes the tissue radioactivity concentration in (Bq/ml), calculated as the ratio between the image intensity and the image scale factor. ID, in (Bq), denotes the injected dose calculated as the product between actual activity and the dose calibration factor [74][90]. Finally, Wp is the patient weight in kilograms.
In this block, an initial operator input is required to loosely identify the region containing the cancer. The main reason to require such an input is because the FDG distribution is, in general, not limited to the cancer alone. Indeed, FDG enters the cells according to glucose transport mechanism, so that active healthy tissues can show high radio-tracer uptake (e.g. brain, heart, bladder, kidneys, and ureters). Thanks to the user's initial direction, whenever the tumor is situated near these anatomical structures, false positives are avoided. However, user intervention is limited to the selection of an initial region of interest ROI0 (Figure 5.3.1-d) around the cancer region in just one axial slice. As a consequence, proposed method is semi-automatic, at least for what concerns FDG-PET applications, and no further intervention is needed. In more favorable conditions, such as in the case of brain metastases, it is possible to take advantage of the greater sensitivity and specificity of MET radio-tracers in discriminating between healthy versus abnormal tissues. Consequently, in MET-PET studies, a fully automatic method could potentially be implemented [84].

5.3.3 Block 3: User Independent Pre-Segmentation Step

Based on the user input ROI0, the block performs an automatic pre-segmentation step, which computes a user independent ROI. This is achieved by constructing a 3D volume of interest (VOI). The first step is to propagate the 2D user ROI0 to adjacent slices. Starting from the slice selected by the operator, neighboring PET slices are visited both upward and downward.
Every time a new slice (index j) is considered, ROI(j-1) is propagated to obtain ROIj, and the peak SUV value within ROIj is detected (SUV^{j}_{max}). If this exceeds 42% [95] of the SUV^{0}_{max}within the initial slice, then the new slice is added to the volume and the next slice is checked. Otherwise no further slices are visited in the specific direction. Of course, when the new jth slice is added to the volume, the propagated contour ROI(j-1) is not expected to be optimal. For this reason, the region growing (RG) algorithm [104] is used to identify a slice-specific boundary. In particular, the SUV^{j}_{max} voxel is used as target seed for a rough 2D segmentation based on the RG method. Following this strategy, an initial 3D VOI is obtained which will contain the SUV^{j}_{max} for all the slices containing the tumor [105].

The next logical step is to locate the voxel with the maximum uptake within the entire VOI (Figure 5.3.1-e). This maximum value may reside on a different slice than the one initially selected by the user, yet it will belong to the same anomaly.
The rough contour (previously generated by the RG algorithm), residing on the

newly identified slice represents the output of the pre-segmentation step (Figure 5.3.1-g) which will feed the next logical block of the system (Figure 5.3.1-h).
It is worth mentioning that the RG algorithm is only used to obtain a rough estimate of the contour encircling the highest radio-tracer uptake area within SUVmax slice, and not for the final segmentation.

Additionally, it is worth noting that this pre-segmentation process generates an output which is independent to inter-operator variations in the initial user input. As such, the user input does not need to be carefully drawn. A rough inclusion of the anomalous area is sufficient to obtain a- more adaptive, and yet repeatable, initial contour. This more adapted contour is then fed into the next step of the system, as shown in step h) Fig.1, where the high-detail delineation is achieved. An example ROI (delineated on the PET image) by the RG method is shown in Figure 5.3.3.

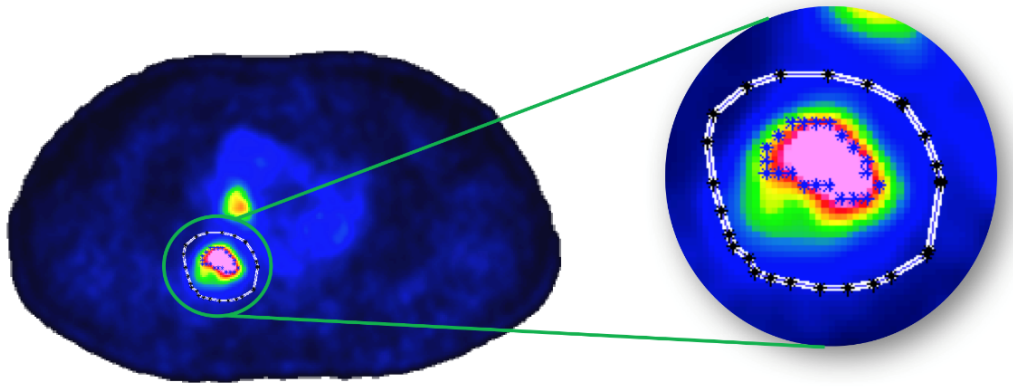


Figure 5.3.3 Operator independent ROI generation on a PET image. The white contour with black asterisks represents the starting ROI provided by the user, while the blue asterisk contour shows the result of the RG algorithm after the slice containing the SUV_{max} is identified. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.3.4 Block 4: Slice Marching Segmentation

This block represents the high detail segmentation obtained through the use of an innovative combination of Local Active Contours (LAC) and KNN classifiers. The LAC algorithm blends the benefits of purely local edge based active contours and fully global region based active contours. At each point along a prominent intensity edge of an object, nearby points inside and outside the object will be modelled well by the mean intensities within the local neighborhoods on either side of the edge. This is the motivation behind the model proposed by Lankton et al [83]; where the

contour energy to be minimized is defined as:

$$E = \oint_C \left(\int_{R_{\text{in}}} \chi_l(x,s)\big(I(x) - u_l(s)\big)^2 dx + \int_{R_{out}} \chi_l(x,s)\big(I(x) - v_l(s)\big)^2 dx \right) ds$$

where Rin and Rout represent the regions inside and outside the curve C, *s* represents the arc length parameter of C, χ represents the characteristic function of the ball of radius l (local neighborhood) centered around a given curve point C(s), I represents the intensity function of the image to be segmented, and $u_l(s)$ and $v_l(s)$ denote the local mean image intensities within the portions of the local neighborhood $\chi_l(x, s)$ inside and outside the curve respectively (within Rin and Rout). These neighborhoods are defined by the function χ, the radius parameter l, and the position of the curve C. Note that the function $\chi_l(x, s)$ evaluates to 1 in a local neighborhood around each contour point C(s) and 0 elsewhere, thereby localizing the processing of image information. The shape of the contour C then divides each such local region into interior and exterior local points in accordance with the contour's role to segment the domain of I. The resulting flow is more robust to initial curve placement and image noise like region-based flows described on [97]; and yet it also capable of capturing significant local structure and partitioning the image without making strong global assumptions about its makeup. To be successful, a segmentation technique based on equation 2 relies on the assumption that nearby points inside and outside the true edge of an object, will be modelled well by the mean intensities within these localized regions. As a result, the energy (2) is more global in nature than edge-based flows but with a "tunable" degree of locality defined by the neighborhood radius l. Actually, the LAC method for 3D MR datasets was applied by Lankton et al [83] via independent segmentation of the 2D slices. A more powerful and coherent segmentation procedure could be performed on all slices simultaneously by evolving a single surface within the corresponding three-dimensional space. While we are currently investigating such 3D shape evolution for future evaluation and publication, the present work moves a step toward 3D data segmentation improving upon Lankton et al [83] by introducing a system to automatically govern the slice-by-slice strategy.

Beyond the optimal identification of the starting slice containing the lesion, and, consequently, the identification of an initial operator independent mask for LAC segmentation, as described previously, further improvements have been introduced in the present algorithm, and in the present block specifically:

- A new energy formulation
- Incorporation of functional information (through SUV)
- An automatic stopping criterion.

These improvements are then integrated into the LAC algorithm, as described below.

First, a new energy formulation is introduced.

In Figure 5.3.1-h, the LAC algorithm receives the optimal mask and the tissue label information from the classifier. Tumor and background regions are not easily separable into two distinct regions. Consequently, energy (2) has been modified to incorporate information from the KNN in form of a new energy term. KNN provides a partition of the PET images into three regions, and this aspect is accommodated as follow:

$$\chi_{lesion}(x) = \begin{cases} 1 & when \;\; KNN(x) = \; lesion; \\ 0 & otherwise; \end{cases}$$

$$\chi_{border-line}(x) = \begin{cases} 1 & when \;\; KNN(x) = \; border - line\,; \\ 0 & otherwise; \end{cases}$$

$$\chi_{\text{background}}(x) = \begin{cases} 1 & when \;\; KNN(x) = \; background; \\ 0 & otherwise; \end{cases}$$

where $\chi_{\text{lesion}}(x)$, $\chi_{\text{border-line}}(x)$ and $\chi_{\text{background}}(x)$ represents the characteristic functions of the tissue's classification respectively for lesion, background, and border-line classes.

The first term of the following formulated energy functional (2) is essentially a prior term penalizing the overlap between regions which are classified in opposite ways by the contour versus the classifier (no penalty is paid in regions classified as "border-line").

To integrate this new prior term and to incorporate SUV in the LAC algorithm, the energy (2) for the PET image segmentation approach is modified to:

$$E = \oint_C \lambda \left(\int_{R_{in}} \chi_l(x,s)\overline{\text{P}}\text{out}_l(x)\,dx \; + \int_{R_{out}} \chi_l(x,s)\overline{\text{P}}\text{in}_l(x)\,dx \right) +$$

$$(1-\lambda)\left(\int_{R_{in}} \chi_l(x,s)(\text{SUV}(x) - u_l(s))^2 dx + \int_{R_{out}} \chi_l(x,s)(\text{SUV}(x) - v_l(s))^2 dx \right) ds$$

where $\lambda \in [0,1]$ is an arbitrary real parameter ($\lambda = 0.01$ provided the best result for our dataset. $\overline{\text{P}}\text{in}_l(x)$ and $\overline{\text{P}}\text{out}_l(x)$ represent the local mean KNN classification within the portions of the local neighborhood inside and outside the curve respectively (within Ω). SUV represents the intensity function of the standardized uptake value to be segmented, $u_l(s)$ and $v_l(s)$ denote the local mean SUV intensities within the portions of the local neighbourhood $\chi_l(x,s)$inside and outside the curve respectively (i.e. within Rin and Rout). These neighborhoods are defined by the function χ, the

radius parameter l, and the position of the curve C. Note that the function $\chi_l(x, s)$ evaluates to 1 in a local neighborhood around each contour point C(s) and 0 elsewhere, thereby localizing the processing of SUV image information. The shape of the contour C then divides each such local region into interior local points and exterior local points in accordance with the contour's segmentation of the SUV. The local means are specified as the ratios of $S_{I_l}(s)$, $S_{E_l}(s)$, $A_{I_l}(s)$, and $A_{E_l}(s)$ which represent the local sums of SUV intensities and the areas of their respective portions of the local neighborhood $\chi_l(x, s)$ inside and outside the curve (within Rin and Rout). More precisely, the local interior region may be expressed as $R_{in} \cap \chi_l(x, s)$ and local exterior region as $R_{out} \cap \chi_l(x, s)$.

$$\overline{\text{Pin}}_l(x) = \frac{\int_\Omega \chi_l(x)\chi_{lesion}(x)\,dx}{\int_\Omega \chi_l(x)\,dx}, \qquad \overline{\text{Pout}}_l(x) = \frac{\int_\Omega \chi_l(x)\chi_{\text{background}}(x)\,dx}{\int_\Omega \chi_l(x)\,dx}$$

$$u_l(s) = \frac{S_{I_l}(s)}{A_{I_l}(s)}, \qquad v_l(s) = \frac{S_{E_l}(s)}{A_{E_l}(s)},$$

$$S_{I_l}(s) = \int_{R_{in}} \chi_l(x, s)SUV(x)\,dx \quad , \quad S_{E_l}(s) = \int_{R_{out}} \chi_l(x, s)SUV(x)\,dx$$

$$A_{I_l}(s) = \int_{R_{in}} \chi_l(x, s)\,dx \quad , \quad A_{E_l}(s) = \int_{R_{out}} \chi_l(x, s)\,dx$$

$$\chi_l(x, s) = \begin{cases} 1 & when \;\; x \in l - Ball(C(s)); \\ 0 & otherwise; \end{cases}$$

After the segmentation is achieved on the slice containing the optimal mask (Figure 5.3.1-h), the contour obtained from the LAC (Figure 5.3.1-i) is used to initiate parallel segmentations on the neighboring slices above and below. Subsequently, for all the following slices (in both directions), we similarly use the segmentation results from the previous slice as initial starting contour. The LAC evolution is inherently capable of locally widening or tightening the contour where necessary. Since, this behavior is driven by the image properties rather than by an inherent knowledge of whether the cancer is present, a stopping criterion is necessary to prevent the

algorithm from misbehaving when there is nothing to be segmented (i.e. tumor is absent in the slice).

Therefore, we devised a fully automatic stopping condition as shown in Figure 5.3.1-k. For the slice under consideration, at each point on the cancer edge, nearby points inside and outside the cancer must have a different local mean SUV. If the cancer is present, a positive difference between background and foreground intensity is expected, and consequently the algorithm can safely proceed with the next neighboring slice (Figure 5.3.1-j). When the system encounters a slice where the local mean vl(S) on Rout is greater or equal to the local mean ul(s) on Rin, which is the opposite of what is expected, the slice is classified as cancer-free and processing of further slices in that direction is halted (Figure 5.3.1-n). The approach can be summarized as:

$$\text{Stop Condition} = \begin{cases} \text{segmentation process is stopped} & when \;\; v_l\,(s) \geq u_l\,(s)\; is\; true; \\ \text{segmentation process is continues} & otherwise; \end{cases}$$

In this way, slice after slice, the BTV is obtained and the segmentation process is automatically stopped, thereby avoiding the need for any user intervention.

5.4 Active contour algorithm with discriminant analysis for delineating tumors in positron emission tomography

In this section, a brief overview of the implemented method is presented (Figure 5.4.1).

In the following sections, the different aspects of the algorithm and its validation will be presented.

The first prerequisite to the use of the algorithm is to convert each PET voxel into SUV to take into account metabolic/functional aspects of the tumor (see following sessions).

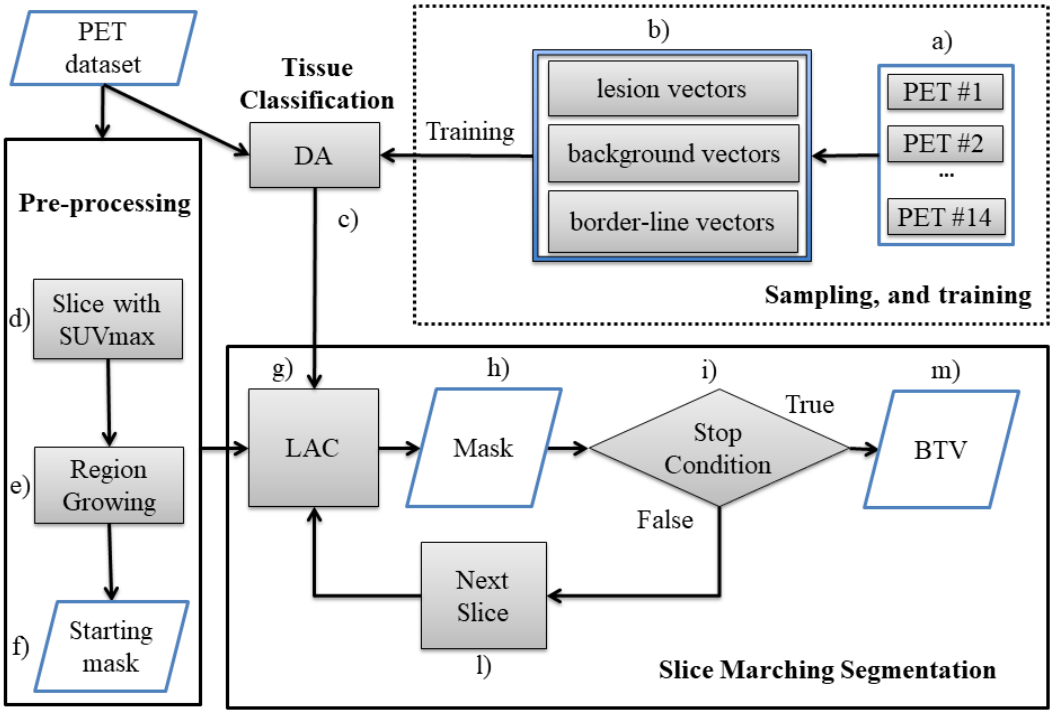


Figure 5.4.1 The flowchart of the proposed segmentation method. a) eight patient lesions, three phantom spheres, and three zeolites are processed by three expert operators; b) sampling operation is obtained for each tissue kind label (lesion, background, and border line); c) training and validation of the DA classifier; d) the lesion area is localized by the operator to avoid any false positive; e) global SUVmax is found in the whole lesion volume; f) RG is used to identify the user independent mask; g) once the initial mask is obtained, it is sent to the next logical block of the system; h) segmentation is performed using the enhanced LAC segmentation algorithm through a new energy based on the DA classification; i) segmentation mask is propagated to the adjacent slices using a slice-by-slice marching approach; l) the stop condition is evaluated; m) segmentation on the next slices is performed until stop condition is false; n) the operator independent BTV is obtained.

To train and validate the DA classifier, tumor, background and border-line regions were automatically identified on PET images using a moving window, 3 by 3 voxels in size, in six phantom experiments and eight patient studies. For each window, the highlighted SUVs were then reorganized in a 9-element vector characterizing the area under investigation (i.e. lesion, background, or border-line tissue). As a result, a total of 4587520 vectors were built for the training and validation of the classifier. The "train and validation" step needs to be performed only once. This one-time operation is highlighted by the dotted line in Figure 5.4.1. Once the above mentioned step has been performed the classifier is able to label new clinical cases and, consequently, to locally drive the segmentation algorithm (Figure 5.4.1c). A thorough explanation of this operation will be given in subsequent sections.

At the beginning of the segmentation process, the region containing the cancer must be localized by the operator. This input ensures that healthy tissues having high radio-tracer uptake or critical conditions are avoided (e.g. when the lesion is located near the heart). Additionally, in case of multiple independent tumors, multiple initial

contours may be setup by the operator so to segment each tumor individually.

The algorithm performs automatically all subsequent operations identifying an optimal starting mask on a PET slice, which can be different from the one initially highlighted by the operator. A more detailed explanation about this task is reported in subsequent sections.

The identified mask is input to the next step of the system, where the segmentation is performed combining a Local region-based Active Contour (LAC) segmentation algorithm appropriately modified to support functional images and the information derived from DA (i.e. tissue classification labels, Fig. 1c) which locally drive the active contour. The obtained segmentation is propagated to the adjacent slices using a slice-by-slice marching approach. Propagation is performed in parallel both upward and downward within the SUV volume to obtain the BTV, until a suitable stopping condition is met. A more detailed explanation about this step is reported in subsequent sections.

5.4.1 Pre-processing of PET dataset

Both the PET studies used to train/validate the DA classifier and the PET studies used to be automatically segmented have been converted into SUV images, as suggested in [86]. To take into account essential functional aspects of the patient, SUV is calculated as the ratio of tissue radioactivity concentration (RC) and injected dose (ID) at the time of injection divided by body weight (Wp):

$$SUV = \frac{RC}{ID/W_p}$$

The RC in (Bq/ml) is calculated as the ratio between the image intensity and the image scale factor. The ID in (Bq) is calculated as the product between actual activity and dose calibration factor [70][90]. Wp is expressed in kilograms.

5.4.2 The discriminant analysis

The DA is used to obtain a tissue labelling to be combined with the image information which locally drive the LAC segmentation. We train DA to split PET tissues into three categories: background, lesion, and border-line regions. The training task needs to be performed only once, and when completed, the DA is ready to classify newly-encountered tissue.

In order to generate the training input to the classifier, three expert operators, with high PET imaging insight and different expertise, manually segmented phantom inserts and tumors from 14 different studies. Such contours, as reported in subsequent sections, were used to obtain the gold standard. To obtain a DA capable

of classifying a wide range of clinical cases, the training comprised two brain metastases, four HNC, two lung cancers, three phantom spheres at different ratio between sphere and background radioactivity concentrations (3:1, 5:1, and 8:1) and diameter (28, 22, and 17 mm, respectively), and three different zeolites (#1, #2, and #6 zeolites described in [89]). Each PET slice containing the target was investigated using a moving window so to extract samples of 3x3 voxels in size. The sample size has been empirically determined to provide the best performance on the present dataset. For each position of the moving window, the selected image portion was compared with the manually produced gold standard. Samples entirely outside the gold standard were labelled as "background". Samples comprising no more than three lesion voxels were labeled as "border-line" tissue. In the remaining cases, the label "lesion" was assigned. Each sample was then reorganized as a 9-element vector. The sampling operation resulted in a total of 4587520 vectors split into 9384 lesion vectors, 5369 border-line vectors, and 4572767 background vectors. The class balancing technique proposed by [91] was performed to assess difference between DA classification using balanced and unbalanced training datasets. This technique proposes a blends under-sampling approach of the majority class with a special form of over-sampling the minority class creating synthetic minority class examples. This study shows that a combination of over-sampling the minority class and under-sampling the majority class can achieve better classifier performance (in Receiver Operating Characteristic – ROC - space) than only under-sampling the majority class. In both cases (balanced and unbalanced training datasets), eighty percent of the samples was used to train the classifier, while the remaining twenty percent was used to check the classifier ability to provide a reliable classification.

The K-Fold cross-validation has been integrated in the classifier to make it reliable and to limit problems such as over-fitting [92][93]. When the training stage is completed, the validation step is started: 917504 vectors are input to the trained DA kernel to verify the performance in classifying new scenarios.

5.4.3 The Initialization Step

We cautiously defined the proposed method as semi-automatic because the user must highlight an initial ROI containing the tumor in order to avoid false positives (i.e. region with high FDG uptake to be considered normal, such as brain, heart, bladder, kidneys, and ureters). Additionally, multiple ROI may need to be defined if independent (i.e. distant and isolated) tumors are present in the same PET. However, no further intervention is needed. The following describes the procedure adopted for the segmentation of one tumor mass.

Based on the user input selection (ROI0), the first step consists of an automatic initialization task which computes a user independent mask. This goal is achieved by first identifying a volume of interest (VOI) which is formed by propagating ROI0 to adjacent slices.

Starting from the slice individuated by the operator, PET slices are visited moving both upward and downward. Every time a new slice (index j) is considered, ROI(j-1) is propagated to obtain ROIj, and the SUVmax within ROIj is detected (SUV_{max}^{j}).
If this value exceeds 42% [94] of SUV_{max}^{0} found within the initial slice, then the new slice is to be added to the volume of interest, otherwise the visitation of further slices in the specific direction is terminated. Of course, when the new (jth) slice is added to the volume, the contour ROI(j-1) is not expected to be optimal. For this reason, the region growing (RG) algorithm [95] is used to identify a slice-specific boundary for each slice. In particular, the SUV_{max}^{j} voxel is used as target seed for the rough 2D segmentation based on the RG. A neighborhood is identified around this voxel to calculate the mean and standard deviation (σ) across a region of connected voxels. The neighborhood is a region with 2r+1 voxels on the side, where r is set to 1. In this way, the neighborhood is a region of three connected voxels forming the side of a rectangle.
The confidence interval intensity range for the seed point is defined as:

$$I(X) \in [\text{mean} - f\sigma, \text{mean} + f\sigma]$$

where I is the image and X is the position of the particular neighbor voxel being considered for inclusion in the region. f is a user provided factor set to 2 [93]. The upper and lower bounds for the eq. 2 are calculated after mean and σ are determined. Neighbor voxels connected to the seed point whose values fall inside this interval are included in the segmented area. The first iteration of the algorithm is completed when no more neighbor voxels are found to fall in the range. During iterations, the mean and σ are updated to determine the new confidence interval.

In this way, a VOI is obtained. Note that by construction, the obtained VOI will contain the SUVmax for all the slices containing the tumor [94]. The initialization is not used for the cancer segmentation, but only to roughly identify the tumor. The next logical step is to locate the voxel with the SUVmax within the VOI, (SUV_{max}^{VOI}). This maximum value may reside on a different slice than the one initially selected by the user but it will belong to the same anomaly. The 2D rough region identified by the RG on the hottest PET slice represents the output of the initialization step.
It is worth noting that the initialization does not require the contour input by the user to be carefully drawn. A rough inclusion of the anomalous area is sufficient. Further, manual drawing does not need to be repeated on any other slice. The benefit of the initialization process is that it generates an output which is independent to local variations in the initial user input, and for that, thereby making the result extremely repeatable. Furthermore, the RG algorithm is used only to obtain a rough estimate of the contour encircling the highest radio-tracer uptake area within the slice containing the SUV_{max}^{VOI}. This contour is then sent to the next step of the system where the delineation is actually carried out using the LAC-DA algorithm which combines an

enhanced LAC segmentation algorithm with information derived from a DA classifier. Further, the LAC-DA is capable to chasing the cancer beyond the VOI boundary (i.e. the final result is not affected/limited by the VOI individuated during the initialization step). An example of the initial mask obtained by the RG during the initialization step is shown in Figure 5.4.2.

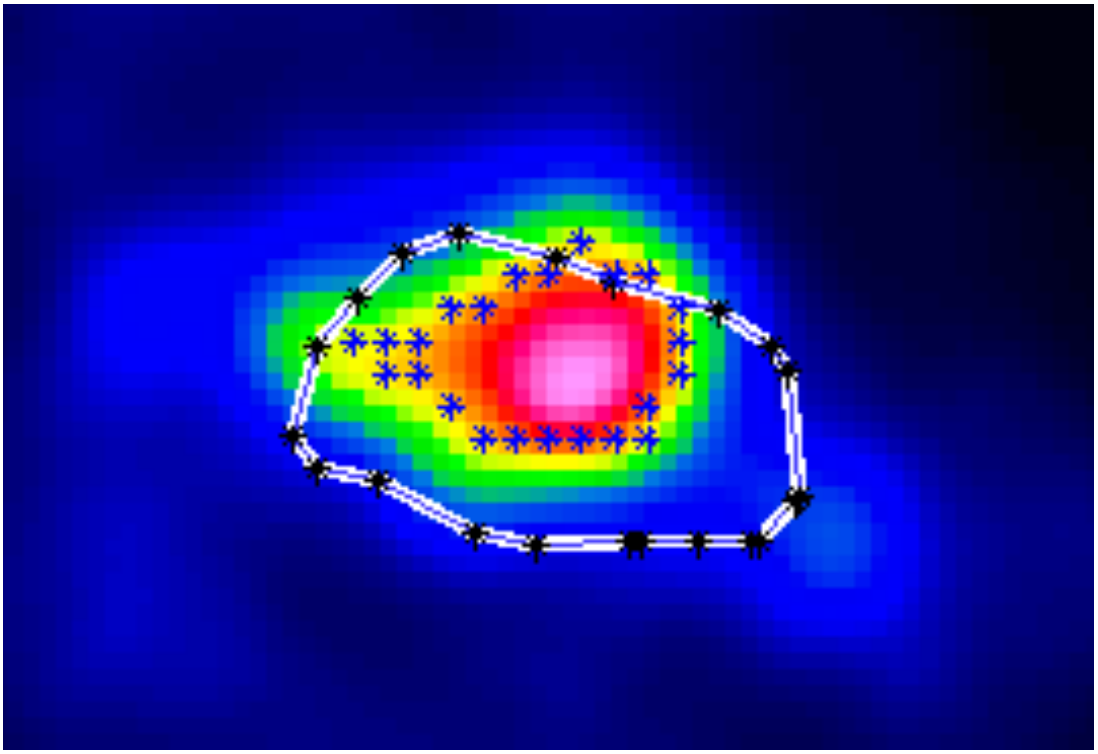


Figure 5.4.2 The white contour with black asterisks represent the rough region highlighted by the user, while the blue asterisk contour shows the result of the initialization step after the slice containing the $\mathbf{SUV_{max}^{VOI}}$ is identified. The latter contour is by construction operator independent and can lie on a slice different from the one initially highlighted by the user.

5.4.4 The LAC-DA algorithm

The LAC algorithm blends the benefits of purely local edge based active contours and fully global region based active contours. At each point along a prominent intensity edge of an object, nearby points inside and outside the object will be modelled well by the mean intensities within the local neighborhoods on either side of the edge. This is the motivation behind the model proposed by Lankton et al [83]; where the contour energy to be minimized is defined as:

$$E = \oint_C \left(\int_{R_{\text{in}}} \chi_l(x,s)\big(I(x) - u_l(s)\big)^2 dx + \int_{R_{out}} \chi_l(x,s)\big(I(x) - v_l(s)\big)^2 dx \right) ds$$

where Rin and Rout represent the regions inside and outside the curve C; where s represents the arc length parameter of C, where χ represents the characteristic function of the ball of radius l (local neighborhood) centered around a given curve point C(s), where I represents the intensity function of the image to be segmented, and where $u_l(s)$ and $v_l(s)$ denote the local mean image intensities within the portions of the local neighborhood $\chi_l(x, s)$ inside and outside the curve respectively (within Rin and Rout). These neighborhoods are defined by the function χ, the radius parameter l, and the position of the curve C. Note that the function $\chi_l(x, s)$ evaluates to 1 in a local neighborhood around each contour point C(s) and 0 elsewhere, thereby localizing the processing of image information. The shape of the contour C then divides each such local region into interior and exterior local points in accordance with the contour's role to segment the domain of I. The resulting flow is more robust to initial curve placement and image noise like region-based flows described on [96]; and yet it capable of capturing significant local structure and partitioning the image without making strong global assumptions about its makeup. To be successful, the present segmentation technique relies on the assumption that nearby points inside and outside the true edge of an object, will be modelled well by the mean intensities within these localized regions. As a result, the energy (2) is more global in nature than edge-based flows but with a "tunable" degree of locality defined by the neighborhood radius l. Actually, the LAC method for 3D MR datasets was applied by Lankton et al [83] via independent segmentation of the 2D slices. A more powerful and coherent segmentation procedure could be performed on all slices simultaneously by evolving a single surface within the corresponding three-dimensional space. While we are currently investigating such 3D shape evolution for future evaluation and publication, the present work moves a step toward 3D data segmentation improving upon Lankton et al [83] by introducing a system to automatically govern the slice-by-slice strategy.

Beyond the optimal identification of the starting slice containing the lesion, and, consequently, the identification of an initial operator independent mask for LAC segmentation, as described in the previous section, further improvements have been introduced in the LAC algorithm:

- A new energy
- SUV functional information
- A fully automatic stopping condition.

These improvements are then integrated into the LAC algorithm, as described below. First, a new energy is introduced.

The LAC algorithm receives the initial user independent mask and the tissue label

information from the classifier. Considering that tumor and background regions are not easily separable into two distinct regions, the energy (2) has been modified to include a new energy term to separate the PET image into three regions considering DA classification as follow:

$$\chi_{lesion}(x) = \begin{cases} 1 & when \;\; DA(x) = \; lesion; \\ 0 & otherwise; \end{cases}$$

$$\chi_{border-line}\,(x) = \begin{cases} 1 & when \;\; DA(x) = \; border - line\,; \\ 0 & otherwise; \end{cases}$$

$$\chi_{\mathrm{background}}(x) = \begin{cases} 1 & when \;\; DA(x) = \; background; \\ 0 & otherwise; \end{cases}$$

where $\chi_{\mathrm{lesion}}(\mathrm{x})$, $\chi_{\mathrm{border-line}}(\mathrm{x})$ and $\chi_{\mathrm{background}}(\mathrm{x})$ represent the characteristic functions of the tissue's classification respectively for lesion, background, and border-line tissues.

The first term of the formulated energy functional (2) is essentially a prior term penalizing the overlap between regions which are classified in opposite ways by the contour versus the classifier (no penalty is paid in regions classified as "border-line").

To integrate this new prior term and to incorporate SUV in the LAC algorithm, the energy (2) for the PET image segmentation approach is defined as:

$$E = \oint_C \lambda \left(\int_{R_{in}} \chi_l(x,s)\overline{\mathrm{P}}\mathrm{out}_l(x)\,dx \; + \int_{R_{out}} \chi_l(x,s)\overline{\mathrm{P}}\mathrm{in}_l(x)\,dx \right) +$$

$$(1-\lambda)\left(\int_{R_{in}} \chi_l(x,s)(\mathrm{SUV}(x) - u_l(s))^2 dx + \int_{R_{out}} \chi_l(x,s)(\mathrm{SUV}(x) - v_l(s))^2 dx \right) ds$$

where the parameter $\lambda \in \mathrm{R}^+$ (range between 0 and 1) is chosen subjectively (in our study λ equal to 0.01 provided the best result). $\overline{\mathrm{P}}\mathrm{in}_{\mathrm{l}}(\mathrm{x})$ and $\overline{\mathrm{P}}\mathrm{out}_{\mathrm{l}}(\mathrm{x})$ denote the local mean DA classification within the portions of the local neighborhood $\chi_{\mathrm{l}}(\mathrm{x})$ inside and outside the curve respectively (within Ω). SUV represents the intensity function

of the standardized uptake value to be segmented, and $u_l(s)$ and $v_l(s)$ denote the local mean SUV intensities within the portions of the local neighborhood $\chi_l(x, s)$ inside and outside the curve respectively (within Rin and Rout). These neighborhoods are defined by the function χ, the radius parameter l, and the position of the curve C.

In our preliminary experiments, the radius parameter l has been determined to provide the best performance in our dataset using trial and error strategy. When using energies based on local features, the radius defining the local neighborhood defines how sensitive to local features the segmentation is meant to be. By tuning this parameter, one can choose the degree to which local and global behavior are weighted. One aspect that must be considered is the order of magnitude of the tumor size and the amount of surrounding noise. When attempting to capture lesions that are very small with nearby noise, as is the case in our dataset, a small localization radius is advisable. For this reason, radius parameter ranged from 3 (very local, small compared to tumor size) to 9 (more global, size large when compared to the features of the tumor edge). After testing, the result suggests a small degradation of the result as the radius increases (i.e. as the energy becomes more global). On the other hand, segmentations were only slightly influenced by the radius. We therefore selected 3 as the most suitable value for the data at hand.

In addition, note that the function $\chi_l(x, s)$ assigns 1 in a local neighborhood around each contour point C(s) and 0 elsewhere, thereby localizing the processing of SUV image information. The shape of the contour C then divides each such local region into interior local points and exterior local points in accordance with the contour's segmentation of the SUV. The local means are specified as the ratios of $S_{I_l}(s)$, $S_{E_l}(s)$, $A_{I_l}(s)$, and $A_{E_l}(s)$ which represent the local sums of SUV intensities and the areas of their respective portions of the local neighborhood $\chi_l(x, s)$ inside and outside the curve (within Rin and Rout). More precisely, the local interior region may be expressed as $R_{in} \cap \chi_l(x, s)$ and local exterior region as $R_{out} \cap \chi_l(x, s)$.

$$\overline{\mathrm{P}}\mathrm{in}_l(x) = \frac{\int_\Omega \chi_l(x)\chi_{lesion}(x)\,dx}{\int_\Omega \chi_l(x)\,dx}, \qquad \overline{\mathrm{P}}\mathrm{out}_l(x) = \frac{\int_\Omega \chi_l(x)\chi_{\mathrm{background}}(x)\,dx}{\int_\Omega \chi_l(x)\,dx}$$

$$u_l(s) = \frac{S_{I_l}(s)}{A_{I_l}(s)}, \qquad v_l(s) = \frac{S_{E_l}(s)}{A_{E_l}(s)},$$

$$S_{I_l}(s) = \int_{R_{in}} \chi_l(x,s) SUV(x)\, dx \quad , \quad S_{E_l}(s) = \int_{R_{out}} \chi_l(x,s) SUV(x)\ \ dx$$

$$A_{I_l}(s) = \int_{R_{in}} \chi_l(x,s)\ \ dx \quad , \quad A_{E_l}(s) = \int_{R_{out}} \chi_l(x,s)\ \ dx$$

$$\chi_l(x,s) = \begin{cases} 1 & when\ \ x \in l - Ball\big(C(s)\big); \\ 0 & otherwise; \end{cases}$$

Given the low resolution relative to tumor size and the diffuse nature of PET images (as well as the SUV images used in our model), the use of a curvature regularizer, which is usually employed to maintain a smooth appearance of the contour, actually revealed unnecessary and therefore is not discussed. Nevertheless, such regularization may become necessary to handle data at higher resolution. As such, the comparative evaluation we performed is based on the two external energies only.

After the segmentation step is achieved on the slice containing SUVmax, the resulting segmented mask is used to initiate parallel segmentations on the neighboring slices above and below. Subsequently, for all the other slices in both directions, we similarly use the segmentation results of the previous slices as the initial mask inputs. The LAC method is inherently capable of locally widening or tightening where necessary when the contour is propagated from slice to slice. Since, this behavior is driven by the image properties rather than by an inherent knowledge of whether the cancer is present, a stopping criterion is necessary to prevent the LAC algorithm from misbehaving, or even diverging, when it reaches a slice where the cancer is absent (i.e. when there is nothing to be segmented).

Therefore, we devised a fully automatic stopping condition. For the slice under consideration, at each point on the cancer edge, nearby points inside and outside the cancer must have a different local mean SUV. If the cancer is present, a positive difference between background and foreground intensity must occur, and consequently the algorithm can safely proceed with the next neighboring slice. When the system encounters a slice where the local mean $v_l(s)$ on Rout is greater or equal to the local mean SUV $u_l(s)$ on Rin, which is the opposite of what is expected, the

slice is classified as cancer-free and the slice-to-slice segmentation propagation is halted in that direction. The approach can be summarized as:

$$\text{Stop Condition} = \begin{cases} \text{segmentation process is stopped} & when \;\; v_l\,(s) \geq u_l\,(s)\; is\; true; \\ \text{segmentation process continues} & otherwise; \end{cases}$$

In this way, slice after slice, the BTV is obtained and the segmentation process is automatically stopped, thereby avoiding the need for any user intervention.

The energy functional used during the contour evolution process is symmetric regarding the "direction of contrast" between the interior and exterior of the evolving contour. In short, along each 2D slice, the contour is equally happy to group darker regions within its interior as it is to group brighter regions within its interior. In a 3D extension of this model, where all slices are segmented simultaneously by a single evolving 3D surface, it would become very sensible to break this symmetry by favoring brighter regions inside the evolving surface and penalizing scenarios where the contrast direction is reversed. In 2D, this ambivalence in the slice-by-slice behavior is exploited to determine the automatic stopping condition. In this way, we don't impose the asymmetry at the level of the 2D active contour. In addition, the direction of contrast established by the initial contour is usually preserved for energies that are symmetric to contrast direction during the evolution of active contours. In the active contour model proposed by [96], the image force acting on the evolving contour along the inward normal direction can be algebraically rewritten as $2(u - v)\{I-(u + v)/2\}$ where u and v are the means inside and outside the curve respectively.

If we consider the effect of this force along (in the absence of regularizing forces based on curvature) and assume be start out with $u > v$, then the sign of the spatially constant scale factor $2(u - v)$ is positive, and the evolution direction will therefore depend point-wise upon whether the image value I is above or below the half-way image intensity threshold of $(u + v)/2$. Contour points where I is above the threshold will move inward, therefore kicking the larger value of I outside the curve while contour points where I is below the threshold will move outward, therefore bringing the smaller values of I inside the curve. This behavior causes u to increase and v to decrease, thereby reinforcing the original direction of contrast. A similar property in our modification of the LAC model allows us to exploit the desired (asymmetric) direction of contrast during the initialization stage of our algorithm.

In the following section, the criteria to judge the quality of the segmentation results are outlined.

5.5 A Fully Automated Segmentation System of Positron Emission Tomography Studies

The main subject of the present study is a fully automatic and operator independent system for brain cancer segmentation, to be used in radiotherapy treatments. While the following subsections will illustrate different components of the system and their validation, in this section we present a brief overview of the system design. The data hereby discussed comprise a total of four phantom experiments and ten oncological patients (see experimental results section). Data from phantoms were used to assess the performances of the delineation algorithm. Concerning the practical use of the system on clinical cases, in order to normalize the voxel activity and to take into account the functional aspects of the disease, the PET images were pre-processed into SUV images (in subsequent sections). The first step is the automatic identification of the optimal combination of starting ROI and slice containing the tumor. Then, this information is input to the subsequent components of the system (in subsequent sections). Once the ROI is identified, the corresponding mask is fed into the next step of the system, where the segmentation is performed combining a Local region-based Active Contour (LAC) algorithm, appropriately adapted to handle PET images. The resulting segmentation is then propagated to the adjacent slices using a slice-by-slice marching approach. Each time convergence criteria are met for a specific slice, the corresponding optimal contour is propagated to the next, where the evolution is continued. Starting from the initial slice, the propagation is performed by contemporarily sweeping the data volume both upward and downward, until a suitable stopping condition, designed to detect a tumor-free slice, is met. Finally, the algorithm outputs a user independent Biological Tumor Volume (BTV). Detailed explanation of this task is provided in subsequent sections.

5.5.1 Pre-processing of PET dataset

Pre-processing PET images is mandatory for inter-patient and follow-up comparisons. Among PET quantification parameters, body-weight SUV is the most widely used in clinical routine. For this reason, it was embedded in our system. SUV is the ratio of tissue radioactivity concentration (RC) and injected dose (ID) at the time of injection, divided by body weight. The RC is calculated as the ratio between the image intensity and the image scale factor. The ID is calculated as the product between actual activity and dose calibration factor.

5.5.2 Interesting Uptake Region identification

In order to obtain a fully automatic BTV segmentation, an initial ROI enclosing the

tumor must be produced, obviously without any intervention by the operator. Therefore, the system identifies the PET slice containing the maximum SUV (SUVmax) in the whole PET volume. By taking advantage of the great sensitivity and specificity of MET radio-tracers in discriminating between healthy and tumor tissues, we can confidently assume that such SUVmax resides inside the main lesion [84].

While this process takes place, an additional test is performed, in order to investigate the presence of isolated local maxima which may indicate metastases separated from the main lesion.

In the case that the presence of multiple (say "n") independent anomalies are recognized, each one is independently processed. A different local maximum (SUV_{max}^{j}, with j = 1:n) is identified for each lesion and, consequently, n regions are automatically identified. By design, the first identified lesion contains the global SUVmax (i.e, SUV_{max}^{1}).

Once the current slice with SUV_{max}^{j} has been identified, an automatic procedure to identify the corresponding ROI starts. The SUV_{max}^{j} voxel is used as target seed for a rough 2D segmentation based on the region growing (RG) method [106]. For each lesion, the obtained ROI represents the output of this preliminary step that is input to the next component of the system, where the actual delineation takes place. The latter is performed through an enhanced LAC segmentation algorithm. It is worth noting that the RG algorithm is used only to obtain a rough estimate of the tumor contour(s).

The same workflow is used to segment each metastasis independently, and the process is designed to carry on automatically. However, in case of multiple lesions the user will receive a warning message and if necessary, will be able to override the default behavior. In such a case, the algorithm can be paused, while the operator inspects the multiple metastases.

5.5.3 The enhanced local active contour method

The model proposed by Lankton et al [83] benefits of purely local edge based active contours and fully global region based active contours. At each point along a prominent intensity edge of the target, nearby points inside and outside the target will be modelled well by the mean intensities within the local neighborhoods on either side of the edge. The contour energy to be minimized is defined as:

$$E = \oint_C \left(\int_{R_{in}} \chi_l(x,s)\big(I(x) - u_l(s)\big)^2 dx + \int_{R_{out}} \chi_l(x,s)\big(I(x) - v_l(s)\big)^2 dx \right) ds$$

Rin and Rout represent the regions inside and outside the curve C, s represents the arc length parameter of C, χ represents the characteristic function of the ball of radius l (local neighborhood) centered around a given curve point C(s), I represents the intensity function of the image to be segmented, $u_l(s)$ and $v_l(s)$ denote the local

mean image intensities within the portions of the local neighborhood $\chi_l(x, s)$ inside and outside the curve respectively (within Rin and Rout).

These neighborhoods are defined by the function χ, the radius parameter l, and the position of the curve C. Note that the function $\chi_l(x, s)$ evaluates to 1 in a local neighborhood around each contour point C(s) and 0 elsewhere. The contour C then divides each such local region into interior and exterior local pixels in accordance with the contour's rule to segment the domain of I.

Beyond the optimal identification of the starting slice containing the lesion, and, consequently, the identification of an initial operator independent mask for LAC segmentation (seen previous sections) further improvements have been introduced in the LAC algorithm. In the following we summarize part of the method, described in [98] . To incorporate metabolic information, the intensity function I in (1) is replaced by the SUV, and $u_l(s)$ and $v_l(s)$ denote the local mean SUV intensities within the portions of the local neighborhood $\chi_l(x, s)$ inside and outside the curve. The shape of the contour C then divides each such local region into interior local points and exterior local points in accordance with the contour's segmentation of the SUV. The local means are specified as the ratios of $S_{I_l}(s)$, $S_{E_l}(s)$, $A_{I_l}(s)$, and $A_{E_l}(s)$ which represent the local sums of SUV intensities and the areas of their respective portions of the local neighborhood $\chi_l(x, s)$ inside and outside the curve. More precisely, the local interior region may be expressed as $R_{in} \cap \chi_l(x, s)$ and local exterior region as $R_{out} \cap \chi_l(x, s)$.

$$u_l(s) = \frac{S_{I_l}(s)}{A_{I_l}(s)} \quad , \quad v_l(s) = \frac{S_{E_l}(s)}{A_{E_l}(s)}$$

$$S_{I_l}(s) = \int_{R_{in}} \chi_l(x, s) SUV(x)\, dx \quad , \quad S_{E_l}(s) = \int_{R_{out}} \chi_l(x, s) SUV(x) \quad dx$$

$$A_{I_l}(s) = \int_{R_{in}} \chi_l(x, s) \; dx \quad , \quad A_{E_l}(s) = \int_{R_{out}} \chi_l(x, s) \; dx$$

$$\chi_l(x, s) = \begin{cases} 1 & \text{when } x \in l - \text{Ball}\big(C(s)\big); \\ 0 & \text{otherwise;} \end{cases}$$

Once the ROI encircling the highest radio-tracer uptake area has been automatically identified (previous section), the resulting mask is used to initiate parallel segmentations on the neighboring slices above and below. Subsequently, for all the other slices in both directions, we similarly use the segmentation results of the previous slices as the initial mask inputs. The LAC method is inherently capable of

locally widening or tightening where necessary when the contour is propagated from slice to slice. Since, this behavior is driven by the image properties rather than by an inherent knowledge of whether the cancer is present, a stopping criterion is necessary to prevent the LAC algorithm from misbehaving, or even diverging, when it reaches a slice where the cancer is absent (i.e. when there is nothing to be segmented).

Therefore, a fully automatic stopping condition is implemented. For the slice under consideration, at each point on the cancer edge, nearby points inside and outside the cancer must have a different local mean SUV. If the cancer is present, a positive difference between background and foreground intensity must occur, and consequently the algorithm can safely proceed with the next neighboring slice. When the system encounters a slice where the local mean $v_l(s)$ on Rout is greater or equal to the local mean SUV $u_l(s)$ on Rin, which is the opposite of what is expected, the slice is recognized as cancer-free and the slice-to-slice propagation is terminated in that direction. In this way, one slice at a time, the BTV is generated. Finally, the segmentation process is automatically stopped, thereby avoiding the need for any user intervention.

5.6 Tissue Classification to Support Local Active Delineation of Brain Tumors

In this section we present a brief overview of the employed segmentation algorithm (Figure 5.6.1). Data from seventeen oncological patients were used partly to train the classifiers, and partly to assess the performances of the delineation algorithm.

Classifiers are used to classify PET voxels according to three labels: tumor, background and border-line. Training and validation were accomplished using a moving window, 3 by 3 voxels, in the PET slices of a data subset. Each window was then reorganized in a 9-element sample vector. The obtained samples were generated and used for training and validation purposes, as detailed in subsequent sections. Training and validation steps are required to be performed only once. After, the classifier can be reused on any new dataset. Concerning the practical use of the system on clinical cases, the first step is the automatic identification of the optimal combination of starting ROI and slice containing the tumor to input the subsequent logical steps of the system (see following sections).

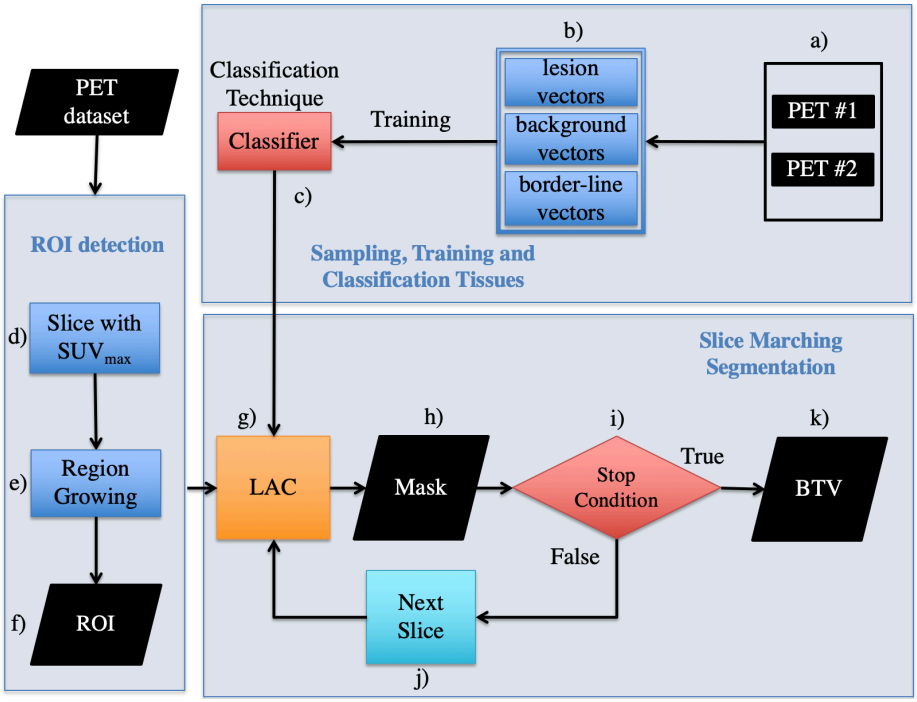


Figure 5.6.1 The segmentation approach [83][101][106]. a) part of the dataset is used to train the classifier; b) samples are obtained for each tissue kind label; c) training and validation of the classifier; d) the region containing the lesion is automatically localized; e) RG is used to identify the user independent mask; f) once the initial mask is obtained, it is sent to the next logical block of the system; g) segmentation is performed using the LAC segmentation algorithm through a new energy based on the tissue classification; h) segmentation mask is propagated to the adjacent slices using a slice-by-slice marching approach; i) a stop condition is evaluated; j) segmentation on the next slices is performed until stop condition is false; k) an operator independent BTV is finally obtained.

Once the ROI is identified, the corresponding mask is feed into the next step of the system, where the segmentation is performed combining a Local region-based Active Contour (LAC) algorithm, appropriately modified to support PET images, and the information derived from the classifiers, which locally drive the active contour (see following sections). The obtained segmentation is then propagated to the adjacent slices using a slice-by-slice marching approach. Each time convergence criteria are met for a slice, the obtained optimal contour is propagated to the next slice, and evolution is resumed. Starting from the first slice considered, propagation is performed in parallel both upward and downward within the data volume and it is continued until a suitable stopping condition, designed to detect tumor-free slice, is met. Finally, a user independent BTV is obtained.

5.6.1 Sampling, Training and Performance of classifiers

In order to normalize the voxel activity and to take into account the functional aspects of the disease, the PET images are pre-processed into SUV images [106].

The classifier (either NB, KNN, or DA) are used to partition PET tissues into three labels: “normal”, “abnormal”, and “border-line” regions. The goal is to combine each classification with the PET image information which locally drive the LAC delineation. Before integrating the classifier in the LAC method, training and validation phases are required to provide the capability of efficiently classify a newly-encountered tissue into the afore mentioned classes. However, the task requires to be performed only once. Once trained, the classifier is ready to be used on any new case. In order to generate the training input to the classifier, two brain metastases, for a total of 16 PET slices, were used (Figure 5.6.1-a). Each PET image containing the lesion is investigated using a moving window of 3x3 voxels in size. The ROI size was empirically determined to provide the best performance on the present dataset. For each new position of the moving window, the selected portion of data was compared with the gold standard. Windows entirely outside or inside the gold standard were labelled as normal or abnormal respectively. Windows comprising no more than three lesion voxels were labeled as border-line tissue. Each window was then reorganized as a 9-element vector (Figure 5.6.1-b). The sampling operation produced a total of 834095 samples; 1706 labeled as lesion vectors, 976 border-line, and 831412 labelled as background. The processing time for single slice was: 54.03 sec for NB, 40.48 sec for KNN, and 26.95 sec for DA (iMac computer with a 3.5 GHz Intel Core i7 processor, 16 GB 1600 MHz DDR3 memory, and OS X El Capitan.). This task is performed only once. The 80% of samples was used to train the three classifiers, while the remaining 20% was used to verify the reliability in classifying newly encountered samples (Figure 5.6.1-c). The K-Fold cross-validation was integrated in the classifiers to make it reliable and to limit problems such as over-fitting. Once the training step was completed, a validation task was performed to verify the rate of success in classifying new scenarios.

5.6.2 The fully automatic protocol

By taking advantage of the great sensitivity and specificity of MET in the discrimination between healthy and tumor tissues in the brain district, the proposed protocol identifies the PET slice containing the maximum SUV (SUVmax) in the whole PET volume to identify the ROI enclosing the tumor without any operator intervention [84]. Once the current slice with SUV_{max}^{j} has been identified, an automatic procedure to identify the corresponding ROI starts. The SUV_{max}^{j} voxel is used as target seed for a rough 2D segmentation based on the region growing (RG) method [106] (Figure 5.6.1-e). For each lesion, the obtained ROI represents the output of this preliminary step (Figure 5.6.1-f) which is feed into the next logical block of the segmentation algorithm (Figure 5.6.1-g) where the actual delineation takes place through the LAC approach. It is worth noting that the RG algorithm is used only to obtain a rough estimate of the tumor contour(s).

5.6.3 The modified local active contour method

In our previous work, the active contour methodology proposed by Lankton et al [83] was improved. For sake of completeness, we summarize the mathematical development of [96][109] regarding how the tissue classification, which separates the PET image into lesion, background, and border-line tissues can be integrated in classical LAC in order to further improve the segmentation process. Briefly, the contour energy to be minimized is defined as:

$$E = \oint_C \left(\int_{R_{in}} \chi_l(x,s)\big(I(x) - u_l(s)\big)^2 dx + \int_{R_{out}} \chi_l(x,s)\big(I(x) - v_l(s)\big)^2 dx \right) ds \quad (1)$$

- Rin denotes the regions inside the curve C
- Rout denotes the regions outside the curve C
- s denotes the arc length parameter of C
- $\boldsymbol{\chi}$ denotes the characteristic function of the ball of radius l centered around a given curve point C(s)
- I denotes the intensity function of the image to be delineated

$u_l(s)$ and $v_l(s)$ represent the local mean image intensities within the portions of the local neighborhood $\chi_l(x,s)$ inside and outside the curve.

Beyond the identification of the initial ROI, as described in the previous section, the energy (1) has been modified to include a new energy term to separate the PET image into three regions considering tissue classification: $\chi_{lesion}(x)$, $\chi_{border-line}(x)$ and $\chi_{background}(x)$ represent the characteristic functions of the tissue's classification (using KNN or NB or DA) respectively for lesion, background, and border-line tissues.

The first term of the formulated energy functional (1) is essentially a prior term penalizing the overlap between regions which are classified in opposite ways by the contour versus the classifier (no penalty is paid in regions classified as "border-line", for this reason the "$\chi_{border-line}(x)$" classification is not included in the energy).

To integrate this new prior term and to incorporate SUV in the LAC algorithm, the energy for the PET image segmentation approach is defined as:

$$E = \oint_C \lambda \left(\int_{R_{in}} \chi_l(x,s)\overline{P}out_l(x)\, dx + \int_{R_{out}} \chi_l(x,s)\overline{P}in_l(x)\, dx \right)$$

$$+(1-\lambda)\left(\int_{R_{in}} \chi_l(x,s)(SUV(x) - u_l(s))^2 dx + \int_{R_{out}} \chi_l(x,s)(SUV(x) - v_l(s))^2 dx \right) ds$$

where the parameter $\lambda \in R^+$ (range between 0 and 1) is chosen subjectively (in our study λ equal to 0.01 provided the best result). $\overline{P}in_l(x)$ and $\overline{P}out_l(x)$ denote the local mean tissue's classification within the portions of the local neighborhood $\chi_l(x)$ inside

and outside the curve respectively (within Ω):

$$\overline{\mathrm{P}}\mathrm{in}_{\mathrm{l}}(\mathrm{x}) = \frac{\int_{\Omega} \chi_{\mathrm{l}}(\mathrm{x})\chi_{\mathrm{lesion}}(\mathrm{x})\,\mathrm{dx}}{\int_{\Omega} \chi_{\mathrm{l}}(\mathrm{x})\,\mathrm{dx}}\,, \qquad \overline{\mathrm{P}}\mathrm{out}_{\mathrm{l}}(\mathrm{x}) = \frac{\int_{\Omega} \chi_{\mathrm{l}}(\mathrm{x})\chi_{\mathrm{background}}(\mathrm{x})\,\mathrm{dx}}{\int_{\Omega} \chi_{\mathrm{l}}(\mathrm{x})\,\mathrm{dx}}$$

$\mathrm{u}_{\mathrm{l}}(\mathrm{s})$ and $\mathrm{v}_{\mathrm{l}}(\mathrm{s})$ denote the local mean SUVs within the portions of the local neighborhood $\chi_{\mathrm{l}}(\mathrm{x},\mathrm{s})$ inside and outside the curve. These neighborhoods are defined by the function χ, the radius parameter l, and the position of the curve C.
Finally, as explained in [98], the process is automatically stopped avoiding the need for any user intervention.

CHAPTER 6

6 Experimental Results

6.1 MR Breast Image Segmentation

6.1.1 Automatic Multi-seed Detection for MR Breast Image Segmentation

The results reported in this section have been used to produce the following research papers:

- **A.Comelli**, A. Bruno, M. L. Di Vittorio, F. Ienzi, R. Lagalla, S. Vitabile, and E. Ardizzone. Automatic Multi-Seed Detection For MR Breast Image Segmentation. Conference: International Conference on Image Analysis and Processing (ICIAP) 2017, At Catania (Italy), Volume: Vol LNCS 10484, pp. 683-693. DOI: 10.1007/978-3-319-68560-1_63

6.1.1.1 University of Palermo Policlinico Hospital Dataset and Materials

The dataset consists of 18 patients from UPPH. The patient were divided in two groups according to their age: group 1 (25/35 years old, glandular/fibroglandular tissue) and group 2 (45/55 years old, fibrofatty/fatty tissue). A GE signa excite1.5 T HD 23 scanner was used to acquire T1 FSE axial sequences with the following technical parameters: 4 channels coil; TR/TE=525; echo train=2; image slices=40; slice thickness=5 mm; slice gap=0; FOV=160×320; bandwidth= 41.67 Hz; imaging matrix=512×256.

6.1.1.2 Gold Standard

Three medical doctors, one resident and two radiologists, with progressively increasing knowledge level of breast imaging, performed the manual segmentation by using DICOM viewer Osirix [35] and following these criteria: breast parenchyma and cutaneous surface were isolated from external air basing on its lower intensity; lower boundary of breast region was delimited by using pectoral muscle as landmark; lateral bounds were represented by axillary cavities. The radiologists usually do not agree with each other, then the results from several observers are used to define a consolidated reference to compare the inter-observer variance, as in [25] [36].

6.1.1.3 Results and Discussion

The proposed method has been tested on the dataset described in the materials section. The results showed that the proposed method achieves excellent results, as depicted in Table 6.1.1. Performance measures are then calculated regarding correct/ incorrect segmentation.

The following measures are computed: Sensitivity, Specificity, Negative Predictive Value, Precision, Accuracy, and Error scores:

- Sensitivity: It is defined as the percentage of effective positives that are correctly identified as such:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

- Specificity: It is defined as the percentage of effective negatives that are not classified as such:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- Precision: It is defined as, related to reproducibility and repeatability, the degree to which repeated segmentations under unchanged conditions show the same results:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Accuracy: It is defined as the degree of closeness of unsupervised segmentations of a breast to that manual segmentation:

$$\text{Accurancy} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

- Overlap: It is defined as the index to quantify agreement between the unsupervised segmentation and manual segmentation:

$$\text{Overlap} = \frac{TP}{(TP + FP + FN)}$$

Furthermore, our method has been compared with a state of the art approach [30] to evaluate the effectiveness and the accuracy in breast segmentation. The results are depicted in Table 6.1.1.The proposed method shows very encouraging results in terms

of statistical metrics (Sensitivity: 95.22%; Specificity: 80.36%; Precision: 98.05%; Accuracy: 97.76%; Overlap: 77.01%) and execution time (4.23 s for each slice).

Table 6.1.1 Experimental results

	Methods	
Segmentation	Proposed method	Template-based method [6]
Sensitivity Eq. (1)	95,22%	88,47%
Specificity Eq. (2)	80,36%	78,12%
Precision Eq. (3)	98,05%	92,99%
Accuracy Eq. (4)	97,76%	92,54%
Overlap Eq. (5)	77,01%	73,84%

6.2 Machine Learning and Data Analysis in Medical Data

The results reported in this section have been used to produce the following research papers:

- **A.Comelli**, M. C. Terranova, L. Scopelliti, S. Salerno, F. Midiri, G. Lo Re, G. Petrucci, and S. Vitabile. A Kernel Support Vector Machine Based Technique for Crohn's Disease Classification in Human Patients. CISIS 2017: In book: Complex, Intelligent, and Software Intensive Systems, pp.262-273. DOI: 10.1007/978-3-319-61566-0_25
- M.C.Terranova, **A.Comelli**, L.Scopelliti, F.Vernuccio, D.Picone, A.Di Piazza, C.Tudisca, F.Midiri, S.Salerno, R.Lagalla, S.Vitabile and G.Lo Re. A New Semi-Automatic Technique for Crohn's Disease Diagnosis Using Supervised Machine Learning Algorithm: Kernel Support Vector Machine. Radiological Society of North America (RSNA) 2017.
- L.Licari, G.Salamone, S.Campanella, F.Carfì, T.Fontana, N.Falco, R.Tutino, P.De Marco, **A.Comelli**, D.Cerniglia, G.Petrucci, S.Vitabile and G.Gulotta. Use of the KSVM-Based System for the Definition, Validation and Identification of the Incisinal Hernia Reccurence Risk Factors. Journal of Surgery 2018. Impact Factor 0.95.

6.2.1 Use of the KSVM-based system for the definition, validation and identification of the incisional hernia recurrence risk factors

6.2.1.1 Patients and methods

In the period from July 2007 to July 2017, 154 patients were selected and subjected to incisional hernia repair. The patients were subjected to surgical procedure at the

Department of Surgical, Oncological and Oral Sciences - General and Emergency Surgery of the Policlinico P. Giaccone of Palermo, University of Palermo. The surgical operations were conducted under general anaesthesia. Patients received antibiotic prophylaxis when indicated, according to the hospital prophylaxis scheme. Inclusion criteria of the study were single operator case studies and open laparotomy for incisional hernia repair. Kernel Support Vector Machine (KSVM) is the intelligent data analysis and data mining proposed to identify and to verify the risk more significant factors for recurrence of incisional hernia. e analysis was conducted verifying 34 risk factors that are shown in Table 6.2.1.

Table 6.2.1 Risk Factors Verified for Recurrence.

Risk factors	
Age	Chevrel R
Sex	Chevrel S
Type of hernia	Cardiovascular morbidities
Urgency regimen of surgical operation	Gastrointestinal morbidities
Surgical technique adopted	Haepatic morbidities
Anaesthesia	Urinary morbidities
ASA score	Other surgical operations
Height	Previous use of mesh
BMI	Type of mesh
Ethnicity	Dimension of the mesh
Job	Dimension of the defect
Endocrine morbidities	Type of suture used
Pulmonary morbidities	Drainage usage
Chevrel M	Post-surgical SSI
Chevrel N	Seroma formation
Chevrel L	Open surgery
Chevrel W	Technique used

6.2.1.2 University of Palermo Policlinico Hospital Dataset and Results

The dataset is composed of 154 patients, 83 females and 71 males, mean age 60 y.o. and a mean BMI of 31 kg/m2.

The preparation for surgical operation evaluated the respiratory function - through chest X-ray, spirometry and pneumological examination - the cardiac function - through electrocardiography and cardio- logical examination - the metabolic indexes, the American Society of Anesthesiologists (ASA) score and the study of parietal defect by clinical examination supplemented by CT scan and CT dynamic scan. The 3D reconstruction of the abdominal defect also allowed a better pre-operative evaluation of the most complex cases and made it possible to select the most suitable surgical procedure and prosthetic materials.

The post-operative complications investigated were seroma formation and Surgical

Site Infection (SSI).
In 28 cases there was evidence of diabetes mellitus type 2 and obesity in 80 patients.
Regarding cardiovascular disease, 70 cases had arterial hypertension associated in 18 cases also with ischemic heart disease.
Eighteen patients were smokers and 12 were ex- smokers; 10 patients were affected by COPD; 21 were affected by thyroid disorders.
Among the various gastrointestinal disorders the most frequently diagnosed were: 11 jatal hernia, 7 GERD, 6 sigmoid diverticulosis and 7 intestinal obstruction.
Among the previous interventions to which patients have been subjected, the most commonly reported were: cholecystectomy in 16 cases, hysterectomy in 11 cases, colectomy in 11 cases, sigmoidectomy in 10 cases, appendectomy in 9 cases.
The localization of the fascial defects was xifopubic in 88 patients, umbilical in 21 patients, mesogastric in 2 cases, hypogastric in 5 cases, umbilical-pubic in 16 cases, pararectal in 2 cases, 4 in trocar, 2 Spigelian hernia and subcostal in 8 cases. Of these cases, 44 were urgent and 110 were with no signs of complications. In 14 cases the diagnosis was of "swiss cheese incisional hernia". The average size of wall defects was 12 cm.
In 12 patients it was possible to perform a direct suture of the defect; in the remaining cases a mesh was positioned to correct the wall defect. Of these, 87 were positioned with IPOM technique, 4 retromuscular (Rives-Stoppa) and 2 preperitoneal.
The general anaesthesia was performed in 141 cases and 109 patients had an ASA score ≥ 3. The re- maining 13 cases were performed in local anaesthesia.
Aspiration drains were placed upon the prosthesis and, if a broad dissection of subcutaneous tissue occurred, also in the subcutis. The drainage removal was performed in 2nd post-operative day or when the drained volume was less than 50 ml/24h.
KSVM was performed to verify the effectiveness of the method. The "Ground-Truth" for incisional hernia is the recurrence. The results showed that the proposed method achieves good performance in terms of correct/incorrect classification of patients.

6.2.1.3 Discussion

Currently, in state of the art, there are not SVM applications on medical data related to incisional hernia. The principal works on data and images classification based on SVM are discussed. We proposed in [127] a new technique for classification of patients affected by Crohn's disease (CD). The proposed technique is based on a Kernel Support Vector Machine (KSVM) and it adopts a Stratified K-Fold Cross- Validation strategy to enhance the KSVM classifier reliability. Son et al. [128] applied an SVM method for data classification. The SVM modelling is a

promising classification approach for predicting medication adherences in Heart Failure (HF) patients. This predictive model helps to stratify the patients so that evidence-based decisions can be made and patients can be managed appropriately. We presented in [93] a system to normal and abnormal tissue classification, using KSVM and k-nearest neighbor (KNN) method, in Positron Emission Tomography Oncological Studies. Zhang et al. [129] proposed a novel hybrid system to classify a given MR brain image as either normal or abnormal.

The goal of the proposed system is to apply these methods to the incisional hernia features extracted for classifying the patient into Positive for recurrence risk or Negative for recurrence risk. The description of algorithms and mathematical formalisms that are used in the proposed classification technique were described. The SVM presents an efficient algorithm and can represent complex non-linear functions. The most popular and most reliable SVMs are the kernel SVMs. Kernel SVMs have the following advantages [130]: work very well in practice and have been remarkably successful in such diverse fields as natural language categorization, bioinformatics and computer vision; have few tunable parameters; training often involves convex quadratic optimization [131].

This study attempts to demonstrate, based on the evidence of scientific literature, the existence of new possible correlations between risk factors and the development of incisional hernia recurrence.

The association between certain characteristics of the patient such as diabetes, chronic lung diseases, smoking, age and increased probability of recurrence was already studied in the guidelines for the prevention of the surgical site infection (SSI) in 1999. The study showed how the contribution of diabetes to the risk of SSI and recurrence is complex. There was a significant relationship between increasing levels of HgA1c and SSI.

In addition, an increase in glucose levels (> 200 mg/dl) in the postoperative period (48h) was associated with an increased risk of SSI. The infection is a risk factor for recurrence too. The infections in diabetics have a double risk.

In DM there is also an alteration of lipid metabolism and a decrease in abdominal wall tissue resistance for ischemic and hypoxic phenomena responsible for wall fragility. The use of nicotine delayed the healing of the wound and may therefore increase the risk of SSI and the development of recurrence [132].

Murariu et al. [133], in 2007, associated age, sex, obesity, pulmonary diseases and diabetes to the development of recurrence.

This data is related to the reduction of reticulin fibers and hyaline degeneration of collagen fibers. There was a recurrence 4 times higher in women (37.3%) especially if obese and with multiple pregnancies. Various mechanical factors have been associated with the development of recurrence including: prolonged abdominal distension in postoperative course (21%), probably because it is responsible for increased abdominal pressure and cough (9%) that creates abdominal hyper- tension by exercising a strong contraction of the diaphragm and abdominal muscles.

Muresan et al. [132] conducted a study in 2016 comparing the various abdominal plastic procedures and found an increasing of recurrence in onlay mesh positioning, especially if associated with changes in abdominal pressure and high post-surgical pain.

In 2017 Hauters et al. [134] demonstrated a significant recurrence increasing in subjects with BMI> 35 (21%), defect greater than 4 cm (27%), overlap <5 cm (32%), and mesh area/defect area <12 (48%). To understand better the pathophysiology of incisional hernia recurrence and in accordance with the evidences emerged from the previous studies, it was decided to investigate new possible criteria involved in the occurrence of the recurrence.

A sample of 154 patients was enrolled in the study. The correlation with the type of surgical technique, the type of prosthesis, its positioning and the post- operative complications were also evaluated.

The data analysis confirmed the known correlations showed in the international literature with a greater incidence of comorbidities such as diabetes (37%), dyslipidaemia and hypercholesterolemia with a cumulative incidence of 16%; tobacco smoke - by combining categories smokers and ex-smokers - reach 46%, COPD 16% and hypertension 51%.

Patients with DM have a higher risk of SSI especially in open interventions. The infection is a risk factor independently for the development of this pathology, so the simultaneous presence of DM should increase the risk of recurrence.

However, in our experience, in patients with diabetes mellitus type 2 who developed a recurrence only in 11 cases there was evidence of infection.

The number of smokers with incisional hernia is high as reported by other studies on patients with abdominal hernia. Several pathogenic mechanisms seem to be involved. Smoking and the hypoxia of peripheral tissues caused by it increase the risk of surgical wound infection and its dehiscence, presumably also by reducing the oxidative killing of neutrophils, which constitute a critical defence against pathogens. In the surgical wounds of smokers there is a de- crease in the deposition of collagen, with the reduction in the ratio between collagen I and III, often found in incisional hernia. COPD is likely that increase the risk of disease by the presence of persistent cough that results in violent abdominal contractions with tension development on wounds. The high incidence of benign prostatic hyperplasia may be correlated with the advanced age of subjects but it is also configurable with DM and dyslipidaemia to the complex metabolic syndrome. The sex is also identified as a risk factor for recurrence with major incidence in females. BMI is not very different between the sexes, whereas there is a slight prevalence of infection in males possibly associated with greater cigarette smoking. The study also revealed a probable role in thyroid pathologies, hepatic disease, cirrhosis and steatosis, benign prostatic hyperplasia, and the presence of uterine fibroids in the development of pathology in the study. In our sample there was a high incidence of hyperthyroidism and nodules, the data could be related to the greater presence of women in the study specimen but also to a probable

role of the immune system. Cirrhosis and liver disease were often associated, especially in advanced stages, with ascites leading to an increase intra-abdominal pressure and increased wall tension.
Benign prostatic hyperplasia seems to play a role because the increased prostate volume could increases the pressure in the abdominal cavity; similar mechanism could be attributed to uterine fibroids, whose high incidence in the study can be attributed to the prevalence of female gender and advanced age.
The greater incidence of xifopubic incisional hernia (57%) seems to confirm previous cases in which a higher frequency of recurrence in longitudinal incisions was observed. The greater incidence (34%) of recurrence is shown in defects ranging from 5 to 10 cm. The infection is a fearful complication; it is usually secondary to a suppurative infection of the surgical wound or to the devascularisation necrosis of the wall layers above the prosthesis.
Sometimes contamination may also occur at the time of mesh positioning by contact with skin bacterial flora. In literature, the incidence is 2-10% after laparotomy and 0-2% after laparoscopy. To prevent this complication it is important to be careful in the skin disinfection and minimize the manipulation of the mesh. In the present study we have reported this complication in 30 cases, with a greater prevalence in male subjects (13-44).

Table 6.2.2 Validity for the Relapse's Classification Methodology.

True Positive	46,00%	Sensitivity	86,25%
True Negative	40,67%	Specificity	87,14%
False Positive	6,00%	Positive Predictive Value	88,46%
False Negative	7,33%	Negative Predictive Value	84,72%
		Accuracy	86,67%
		Precision	88,46%
		Error	13,33%

6.2.2 A Kernel Support Vector Machine Based Technique for Crohn's Disease Classification in Human Patients

6.2.2.1 University of Palermo Policlinico Hospital Dataset

The dataset consists of 300 MR Enterography examinations of 300 patients coming from University of Palermo Policlinico Hospital, Department of Radiology, (156 females, 144 males, mean age 37,8 years). MR imaging studies were performed with a 1.5-T magnet and surface coils phased array (Signa, GE Medical system, Milwaukee, WI, USA and Achieva, Philips Medical System, Eindhoven, The

Netherlands), with the use of paramagnetic contrast media (DOTAREM; Guerbet; USA), after the administration of a spasmolytic agent, unless contraindicated.
In our protocol has been used the following sequences: HASTE thick-slab, Steady State Free Precession, Single Shot Fast Spin Echo, 3D Spoiled Gradient Echo - on axial and coronal planes - and DWI. Patients underwent specific protocol for MR Enterography, that requires 6 hours fasting before the exam, low fiber diet for the preceding 5 days, and the intake from the day before the exam of 2000 ml of water and polyethylene glycol (PEG). Thus allowing to reduce the fecal matter and to provide a better and easier distention of the bowel loops. Upon the arrival to the department, patient is invited to ingest other 1500ml of water and PEG 40min before the exam.
The dataset is composed of 300 E-MRI examinations related to 300 patients: 150 are histologically CD proved, and 150 are healthy individuals [43][50][51]. Each patient has been codified with a vector of 22 qualitative features. The expert radiologist extracted features are reported in Table 4.2.1.

6.2.2.2 Evaluation Metrics, Results and Discussion

The proposed technique has been used on the datasets described in the Materials section. Currently, in state of the art, there are not KSVM applications on medical data related to CD. For that reason, it was not possible to propose a method of comparison in addition to ours. To test the effectiveness of the technique, a direct comparison with four SVMs different kernels algorithms (Linear, Polynomial, Quadratic, RBF, reported in Table 4.2.2) has been performed. When using the linear kernel, the KSVM degrades to original linear SVM. It is important to highlight that the Ground-Truth for CD diagnosis is the histological specimen result: it is the absolute certainty. The biopsy is performed under endoscopic examination. Two biopsies from five sites around the colon, including the rectum, are obtained, and then analyzed by expert pathologists of University Hospital Policlinico, for diagnosis confirmation according to European Crohn's and Colitis Organisation's guide-lines [44]. The results showed that the proposed technique obtains excellent results in both training and validation as reported in Table 6.2.3 and in
Table 6.2.4. Performance measures are then calculated regarding correct/incorrect Patient classification:

- True Positive (TP): a vector is considered Positive for CD e both in histological examination and in supervised classification;
- False Positive (FP): a vector is considered Positive for CD in supervised classification and Negative for CD histological examination;
- True Negative(TN):a vector is considered Negative for CD both in

histological examination and in supervised classification;

- False Negative (FN): a vector is considered Positive for CD in supervised classification and Negative for CD histological examination.

At this point it is possible to calculate the Sensitivity, Specificity, Negative Predictive Value, Precision, Accuracy, and Error scores:

- Sensitivity: It is defined as the percentage of effective positives that are correctly identified as such:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

- Specificity: It is defined as the percentage of effective negatives that are not classified as such:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- Negative Predictive Value: It is defined as the probability that subjects with a negative test are truly not diseased:

$$\text{Negative Predictive Value} = \frac{TN}{TN + FN}$$

- Precision: It is defined as, related to reproducibility and repeatability, the degree to which repeated classifications under unchanged conditions show the same results:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Accuracy: It is defined as the degree of closeness of classifications of a desease to that desease's true histological examination:

$$\text{Accurancy} = \frac{TP + TN}{TP + FP + FN + TN}$$

- Error: It is defined as the opposite of Accuracy:

$$\text{Error} = \frac{FP + FN}{TP + FP + FN + TN}$$

The classification result has been compared with three Stratified K-Fold validation, with same SVM Linear Kernel. The results are depicted in Table 6.2.3.

Table 6.2.3 Classification comparison of 3 stratified K-Fold validation strategies using the same SVM Linear Kernel

Classification	Linear SVM Kernel		
	Fold cross validation		
	5	10	15
True Positive	47,00%	46,20%	46,33%
True Negative	50,00%	50,00%	50,00%
False Positive	0,00%	0,00%	0,00%
False Negative	3,00%	3,80%	3,67%
Sensitivity Eq. (1)	94,00%	92,40%	92,67%
Specificity Eq. (2)	100,00%	100,00%	100,00%
Negative Predictive Value Eq. (3)	94,34%	92,94%	93,17%
Precision Eq. (4)	100,00%	100,00%	100,00%
Accuracy Eq. (5)	97,00%	96,20%	96,20%
Error Eq. (6)	3,00%	3,80%	3,80%

For 5-Fold Cross Validation, the whole classification accuracy was 97,00%; for 10-fold Cross Validation, was 96,40%; and for 15-Fold Cross-Validation, was 96,40%. Obviously, the 5-Fold Cross Validation outperformed the other two Strati- fied K Fold validation. The classification result has been compared with four different KSVM, with same 5-Fold Cross Validation. The results are depicted in Table 6.2.4.
For Linear kernel, the whole classification accuracy was 97,00%; for Polynomial kernel, was 96,40%; for the Quadratic kernel, was 96,40%; and for the RBF kernel, was 97,40%. Obviously, the RBF kernel SVM outperformed the other three kernel SVMs.

Table 6.2.4 Classification comparison of 4 different SVM Kernels

Classification	5-Fold cross validation			
	SVMs Kernel			
	Linear	Polynomial	Quadratic	Radial basis function
True Positive	47,00%	46,40%	46,40%	47,40%
True Negative	50,00%	50,00%	50,00%	50,00%
False Positive	0,00%	0,00%	0,00%	0,00%
False Negative	3,00%	3,60%	3,60%	2,60%
Sensitivity Eq. (1)	94,00%	92,80%	92,80%	94,80%
Specificity Eq. (2)	100,00%	100,00%	100,00%	100,00%
Negative Predictive Value Eq. (3)	94,34%	93,28%	93,28%	95,06%
Precision Eq. (4)	100,00%	100,00%	100,00%	100,00%
Accuracy Eq. (5)	97,00%	96,40%	96,40%	97,40%
Error Eq. (6)	3,00%	3,60%	3,60%	2,60%

6.3 Image Segmentation in PET

The results reported in this section have been used to produce the following research paper:

- **A.Comelli**, A. Stefano, V. Benfante, and G. Russo. Normal and Abnormal Tissue Classification in PET Oncological Studies. Pattern Recognition and Image Analysis, 2018, Vol. 28, No. 1, pp. 121–128. DOI: 10.1134/S1054661818010054. Impact Factor: 0.998.
- **A.Comelli**, A. Stefano, G. Russo, M.G. Sabini, M. Ippolito, S. Bignardi, G. Petrucci, A. Yezzi. A smart and operator independent system to delineate tumours in Positron Emission Tomography scans, Computers in Biology and Medicine. Accepted 6 September 2018. DOI: 10.1016/j.compbiomed.2018.09.002. Impact Factor: 2.115.
- G.Russo, **A.Comelli**, A. Stefano, M.G.Sabini, M.Ippolito, M.C.Gilardi, A.Yezzi. Sviluppo di un metodo per la segmentazione del BTV accurato e operatore indipendente. 10°Congresso Nazionale (AIFM) 2018.
- **A.Comelli**, A.Stefano, S.Bignardi , G.Russo, M.G.Sabini, M.Ippolito, S.Barone and A.Yezzi. Active Contour Algorithm with Discriminant Analysis for Delineating Tumors in Positron Emission Tomography. Artificial Intelligence in Medicine. Accepted 7 January 2019. DOI: 10.1016/j.artmed.2019.01.002. Impact Factor: 2.879.
- G.Russo, **A.Comelli**, A.Stefano, M.G.Sabini, M.Ippolito, M.C.Gilardi, & Yezzi. 253. An accurate and operator independent method for biological

tumour volume segmentation. Physica Medica: European Journal of Medical Physics, 56, 218. Accepted December 2018. DOI: 10.1016/j.ejmp.2018.04.264.

- **A.Comelli**, A.Stefano, G.Russo, S.Bignardi, M.G.Sabini, G.Petrucci, M.Ippolito, and A.Yezzi. K-Nearest Neighbor driving Active Contours to Delineate Biological Tumor Volumes. Engineering Applications of Artificial Intelligence. Under review 2018. DOI: 10.1016/j.engappai.2019.02.005.
- **A.Comelli**, A.Stefano, S.Bignardi, C.Coronnello, G.Russo, M.G.Sabini, M.Ippolito, and A.Yezzi. Tissue Classification to Support Local Active Delineation of Brain Tumors. In Annual Conference on Medical Image Understanding and Analysis. Springer, Cham, 2019.
- **A.Comelli**, and A.Stefano. A fully automatic system of Positron Emission Tomography Study segmentation. In Annual Conference on Medical Image Understanding and Analysis. Springer, Cham, 2019.

6.3.1 A smart and operator independent system to delineate tumours in Positron Emission Tomography scans

6.3.1.1 Framework for performance evaluation

A framework for the evaluation of the proposed segmentation system is presented. Overlap-based and spatial distance-based metrics, according to the formulations presented in [101][110][100], are considered to determine the accuracy achieved by the proposed computer-assisted segmentation system (assessed volume) s_{Seg} against the gold-standard (reference volume) s_{ref} (in our study, manual segmentations performed by three experts are used to define a consolidated reference as described in the next Section).

The overlap and difference between the two contours were measured according to true positive (TP), false positive (FP), true negative (TN) and false negative (FN) voxels. In particular, we calculated the mean, standard deviation, 95% confidence interval (CI) and coefficient of variation (CV) of sensitivity, positive predictive value (PPV), dice similarity coefficient (DSC), Hausdorff distance (HD), and Mahalanobis distance (MHD).

Sensitivity, also called the true positive volume fraction (TPVF), is the fraction of the total amount of tissue in the proposed segmentation system s_{Seg} which overlaps with the reference volume s_{ref} [101]. A perfect segmentation algorithm would be 100 % sensitive (segmenting all voxels from the target voxels) and 100 % specific (not segmenting any from the background voxels). In particular, TPVF [99] is a crucial measure in radiotherapy in order to obtain optimal RTP avoiding cancer recurrence.

Specificity, also called the true negative volume fraction (TNVF), is the fraction of tissue defined in the reference volume s_{ref} that is missed by the segmentation system s_{Seg}. Considering that the number of true negatives depends on the space volume I, the specificity makes little sense and only the sensitivity conveys useful information. The specificity can be replaced with the positive predictive value (PPV) [100].
PPV, also called precision, is the fraction of the total amount of tissue in the reference volume s_{ref} which overlaps with the segmentation system s_{Seg}.
DSC value [101] measures the spatial overlap between the reference volume s_{ref} and the segmentation system s_{Seg}: a DSC value equal to 100% indicates a perfect match between two volumetric segmentations, while a DSC whose value is 0% indicates no overlap:

$$DSC = \frac{2|s_{Seg} \cap s_{ref}|}{|s_{Seg}| + |s_{ref}|} = \frac{2\times TP}{2\times TP + FP + FN} \times 100\%$$

Nevertheless, overlap-based metrics are highly dependent on the segmentation size. For this reason, distance-based metrics are highly recommended when the boundary segmentation is critical, such as in BTV delineation for RTP. In particular, HD is used to measure the most mismatched boundary voxels between automatic and manual BTV: a small median of HD means an accurate segmentation, while a large median of HD means no accuracy.
The HD [110] [136] between two finite point sets A and B is defined by:

$$HD = max\{h(A,B), h(B,A)\}$$

The two distances h(A, B) and h(B, A) are termed as forward and backward Hausdorff distances of A to B. h(A, B) identifies the point a∈A that is farthest from any point b∈B and it measures the distance from a to its nearest neighbour in B. h(B, A) identifies the point b∈B that is farthest from any point a∈A and it measures the distance from b to its nearest neighbour in A.
Each point of A (or B) must be within distance h(A,B) of some point of B, and there is at least one of A that is exactly distance h(A,B) from the nearest point of B considered as the most mismatched point. Symmetric consideration for h(B,A), with respect to B instead of A.
h(a,B) is defined as $\min_{b\in B}$ h(a,b) where h(a,b) is the Euclidean distance, and, similarly, h(b, A) is defined as $\min_{a\in A}$ h(b, a). Consequently, h(A, B) is defined as the $\max_{a\in A}$ h(a, B), and h(B, A) is defined as the $\max_{b\in B}$ h(b, A). Then,

$$\mathrm{h(A,B)} = max_{a\in A}\left\{ min_{b\in B} \sqrt{\sum_{k=1}^{n}(a_k - b_k)^2} \right\}$$

where the Euclidean distance is defined on n-dimensional Euclidean space. Similarly for h(B,A). Thus, HD measures the degree of mismatch between A and B by measuring the distance of the point of A that is farthest from any point of B, and vice versa.
Since medical images are usually characterized by noise and the HD is generally sensitive to outliers, the HD averaged over all points in A and B has been considered [110]:

$$HD = max\{d(A,B), d(B,A)\}$$

where $d(A,B) = (1/N)\sum_{\mathbf{a}\in A} \min_{\mathbf{b}\in B} \|\mathbf{a}-\mathbf{b}\|$.

A variant of the MHD has been considered to take into account the correlation of all points belonging to two different points clouds [110]. The MHD between the sets of voxels contained in s_{Seg}e s_{Ref} is given by:

$$\text{MHD} = \sqrt{(\mu_{Seg} - \mu_{Ref})^T S^{-1} (\mu_{Seg} - \mu_{Ref})}$$

where μSeg and μRef are the means of the two segmentations. S is the common covariance of the two sets and is given by

$$S = \frac{n_1 S_{Seg} + n_2 S_{Ref}}{n_1 + n_2}$$

where SSeg, SRef are the covariance matrices of the voxel sets and n1, n2 are the numbers of voxels in each set. In phantom experiments, shape modifications and volume translations are unlikely to be observed. In these studies, sensitivity, PPV, DSC and HD measurements can be considered more than sufficient for performance assessment, as reported in [101]. For this reason, MHD has been considered in only patient studies.
Finally, to perform statistical test between the proposed system and human segmentations, the combination of sensitivity and PPV, as recommended in [78], has been used. Further, to facilitate the evaluation and ranking of the results, three new accuracy scores have been defined:

- Score = 0.5 × sensitivity + 0.5 × PPV;
- Score radiotherapy planning (RT) = 0.6 × sensitivity + 0.4 × PPV;
- Score follow-up (FU)= 0.4 × sensitivity + 0.6 × PPV.

In radiotherapy planning, the aim is to reduce the risk of missing the target, even if it means delivering higher dose to the surrounding healthy tissues and organs-at-risk. For this reason, sensitivity could be considered more important than PPV. For

therapy follow-up the aim is to obtain consistent volume measurements in sequential PET scans and to avoid including background/nearby tissues. As a result, PPV could be considered more important than sensitivity.

The performance of the proposed system is compared to other state of the art PET image segmentation methods. In particular, the fixed thresholding (42% of the SUVmax) method (T42%) [95], the RG method [96], the FCM clustering method [109], the enhanced RW method such as described in [86], and the original LAC method [83] have been used for comparison. The software package used to provide manual gold-standard and proposed segmentation BTVs and evaluation task has been implemented in the Matlab R2016a simulation environment (MathWorks, Natick, MA, USA), running on an iMac computer with a 3.5 GHz Intel Core i7 processor, 16 GB 1600 MHz DDR3 memory, and OS X El Capitan.

6.3.1.2 Phantom Studies

National Electrical Manufacturers Association International Electrotechnical Commission (NEMA IEC) phantom is used for preliminary performance testing [86][27][87]. The phantom is composed of an elliptical cylinder (D1=24 cm, D2=30 cm, h=21 cm) with six different spherical lesions of size 10, 13, 17, 22, 28, and 37 mm in diameter placed at 5.5 cm from the centre of the phantom. Both body phantom and spheres are filled with a known amount of radioactive tracer to simulate oncological lesions. It could be argued that cancer is often inhomogeneous and irregularly shaped in contrast with spherical targets and that tracer-filled spheres suffer from cold-body effects when compared to the real case (the human body) [78][88]. However, the aim of this test is to evaluate the efficiency of the system and fairly compare its performance with other methods under identical conditions where results may be quantitatively and reliably evaluated. For this purpose, using a well-defined sharp-edged target is preferable because the only uncertainty on the exact boundary location is introduced by the point spread function of the measuring system. The ratio between sphere concentrations and background radioactivity concentrations ranged from 1.5:1 to 8:1 for five independent experiments: 1.5:1 for the phantom “I”, 3:1 for the phantom “II”, 5:1 for the phantom “III”, 7:1 for the phantom “IV”, and 8:1 for the phantom “V”. Performance results are calculated considering small spheres (sphere diameter less than or equal to 17 mm) and large spheres (sphere diameter greater than 17mm). This choice was motivated by the fact that large biases are introduced [137][77] when the lesion size is smaller than 2–3 times the Full Width at Half Maximum (FWHM) of the PET point spread function. The PET/CT acquisition protocol is described in the Section 5.

6.3.1.3 Clinical studies

In the present study, we retrospectively considered 50 cases acquired at the Nuclear Medicine Department of the Cannizzaro Hospital (Catania, Italy). The dataset contains 50 PET examinations of patients with lung, head & neck, and brain cancers that had been referred for a diagnostic PET/CT scan before radiotherapy treatments. Segmentations were performed off-line and the results had no influence on the treatment protocol, nor on the patient management. No sensitive patient information was accessed. As such, the institutional Hospital Medical Ethics Review Board approved this study protocol and all subjects involved were properly informed and released their written consent. In FDG PET studies, patients fasted for 12 h before the PET examination, and successively were intravenously injected with FDG. The PET/CT oncological protocol began 60 min after the injection. Patients breathed normally during the PET and CT examinations, and scanning was executed from the top of the skull to the middle of the thigh with the arms along the body. In MET PET studies, for only brain acquisitions, patients fasted for 4 h before the PET examination. The PET/CT protocol began 10 min after the MET injection.

6.3.1.4 PET/CT Acquisition Protocol

The acquisitions for both phantom experiments and clinical studies were performed within the same Nuclear Medicine Department using the same equipment, a Discovery 690 PET/CT scanner with time of flight (General Electric Medical Systems, Milwaukee, WI, USA). The phantom and patient protocols included a SCOUT scan at 40 mA, a CT scan at 140 keV and 150 mA (10 sec), and 3D PET scans (2.5 minutes per bed position). PET images were reconstructed using a 3D ordered subset expectation maximization (OSEM) algorithm. All imaging data were encoded in the 16-bit DICOM format.

Each PET slice consists of 256×256 voxels with a grid spacing of 2.73 mm3 and thickness of 3.27 mm3. Consequently, the size of each voxel is 2.73 × 2.73 × 3.27 mm3. Thanks to the injected PET radiotracer (FDG or MET), tumours appears as hyper-intense regions.

The non-diagnostic CT scan is performed for attenuation correction and anatomic localization of the tumour contextually to PET image acquisition. The CT slice consists of 512 × 512 voxels with size 1.36 × 1.36 × 3.75 mm3.

6.3.1.5 Gold Standard

Phantoms offer the advantage of known target boundaries. Consequently, we used the match between segmented CT images and the known position of the spherical shapes to quantitatively evaluate the performance of the proposed segmentation system.

In patient studies, knowledge of the ground truth would require exact knowledge of the pathology region. In PET imaging, histopathology analysis provides the only valid ground truth for quantitative assessment [100][108]. Since in radiotherapy the histopathology analysis is unavailable, the actual gold-standard is impossible to retrieve. To assess the clinical effectiveness and feasibility of a proposed segmentation method it is therefore standard practice to refer to manual delineations performed by expert professionals as a substitute for ground truth [78]. The BTV delineation is actually a critical task performed by expert physicians to determine which areas to include or exclude in the planned target volume. Yet, manual delineation carries a certain amount of subjectivity and is often influenced by the clinical specialization of the operator. For example, radio-therapists will, on average, draw larger boundaries than oncologists. For this reason, we used as gold-standard the segmentations performed by three experts with high clinical and PET imaging insight and different expertise (the chief nuclear medicine physician –M.I. author-, the chief radiotherapy physician –M.S. author- and an expert radiotherapy physician –G.R. author-).
In order to investigate the inter-observer variance, a simultaneous ground truth estimation tool was employed (as suggested by [36][98]), and the results from three observers were combined to define a consolidated reference.

6.3.1.6 Clinical testing and Results on Phantoms

Performance results have been divided in two tables considering small spheres (sphere diameters: 10, 13, and 17 mm) and large spheres (sphere diameters: 22, 28, and 37 mm) as explained in section 5.
Table 6.3.1 and Table 6.3.2 show sensitivity, PPV, DSC and HD values in five independent phantom experiments carried with different ratios between sphere concentration and background radioactivity concentration: 1.5:1 for the phantom “I”, 3:1 for the phantom “II”, 5:1 for the phantom “III”, 7:1 for the phantom “IV”, and 8:1 for the phantom “V”.

Table 6.3.1 Sensitivity, PPV, DSC, HD rates obtained over NEMA IEC phantoms (II–V). Phantoms have the following signal contrast: (II) 3:1, (III) 5:1, (IV) 7:1, and (V) 8:1. Three different spherical lesions of size 10, 13, and 17 mm in diameter are considered to assess the performance of the proposed segmentation system. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

NEMA IEC Phantom	Sphere Diameter	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]
II	10 mm	66.71%	76.22%	71.15%	1.00
III		67.62%	76.71%	71.96%	1.47
IV		75.81%	73.50%	74.63%	1.00
V		71.10%	82.16%	76.22%	1.11
II	13 mm	66.73%	92.31%	77.40%	1.00
III		75.04%	81.81%	78.38%	1.08
IV		66.01%	97.17%	78.60%	1.00
V		70.03%	91.30%	79.20%	1.00
II	17 mm	66.00%	97.16%	78.61%	1.21
III		69.95%	96.33%	81.06%	1.28
IV		73.43%	92.00%	81.60%	1.20
V		71.61%	94.07%	81.35%	1.07
Mean		70.00%	87.56%	77.51%	1.12
± std		±3.49%	±8.87%	±3.46%	±0.15
± CI		±1.97%	±5.02%	±1.95%	±0.08
± CV		±4.98%	±10.13%	±4.46%	±0.13

Table 6.3.2 Sensitivity, PPV, DSC, HD rates obtained over NEMA IEC phantoms (I–V). Phantoms have the following signal contrast: (I) 1.5:1, (II) 3:1, (III) 5:1, (IV) 7:1, and (V) 8:1. Three different spherical lesions of size 22, 26, and 37 mm in diameter are considered to assess the performance of the proposed segmentation system. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

NEMA IEC Phantom	Sphere Diameter	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]
I	22 mm	72.03%	72.51%	80.52%	1.77
II		91.20%	88.05%	89.60%	1.17
III		92.01%	85.81%	88.82%	1.06
IV		90.52%	86.42%	88.41%	1.07
V		96.33%	87.90%	91.90%	1.00
I	26 mm	89.22%	81.32%	84.87%	1.71
II		92.91%	90.90%	91.90%	1.18
III		91.87%	88.98%	90.31%	1.05
IV		94.40%	90.20%	92.20%	1.00
V		99.80%	85.21%	91.91%	1.05
I	37 mm	89.04%	90.90%	89.90%	1.60
II		99.90%	88.22%	93.72%	1.19
III		99.90%	88.63%	93.90%	1.05
IV		99.41%	91.50%	95.35%	0.97
V		99.82%	93.81%	96.70%	0.89
Mean		93.22%	87.36%	90.67%	1.18
± std		±7.18%	±5.09%	±4.07%	±0.28
± CI (95%)		±3.64%	±2.58%	±2.06%	±0.14
± CV		±7.71%	±5.83%	±4.49%	±0.23

Table 6.3.1 shows the volumetric accuracy results for the smaller spheres, while Table 6.3.2 shows the volumetric accuracy results for the larger spheres. In Table 1, the results for the phantom "I" are missing because the ratio between sphere and background concentrations is too low (the target radioactivity concentration is only more than one-and-a-half times the background radioactivity concentration), and no PET region with high uptake concentration is visible for spheres with diameter < 22 mm. No significant difference between background and target is observed. For the same reason, the accuracy improved for all spheres, regardless of their volume, when the ratio between sphere concentration and background radioactivity concentration was increased. Phantom images with different signal contrast (1.5:1 and 8:1, respectively) are shown in Figure 6.3.1.

In general, due to the PVE, the separation of small targets from the background region is very challenging, and the difficulty increases in critical conditions (i.e. low signal contrast). The volumes of smaller spheres are underestimated (mean difference between segmented and actual volumes = -19.12±10.69%) with more false negatives than false positives. As expected, large errors occur in lesions less than 2 cm in diameter [77].

Table 6.3.2 shows the results for all spheres with a diameter greater than 17 mm. In all conditions, excluding the phantom 'I' for the aforementioned reason, a DSC rate above 90% and a sensitivity rate greater than 90% are observed. These performances are generally accepted as excellent. A slight oversizing is observed (mean difference between segmented and actual volumes = 6.69±5.23%), nevertheless, larger margins can help to prevent the extension of tumour infiltration in radiotherapy treatments [111].

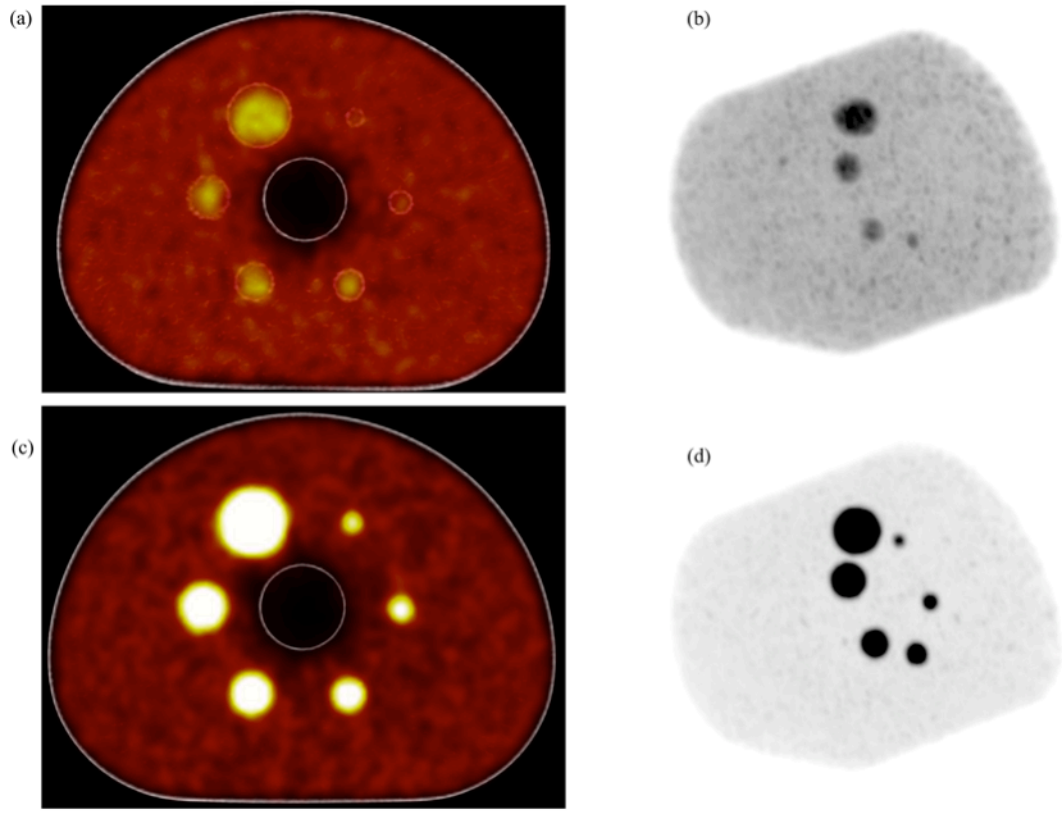


Figure 6.3.1 PET/CT images of I and V phantoms. Phantoms have the following signal contrasts: (a) 1.5:1, (c) 8:1. PET volume rendering of phantom I (b), and V (d).

In addition, high sensitivity, PPV and DSC, and low HD confirm the accuracy of the proposed segmentation system even when compared against the most common segmentation algorithms used in BTV extraction (RW, original LAC, RG, FCM, and T42%). Figure 6.3.2 reports the quantitative comparison between the semi-automatic segmentation and the gold-standard considering all phantom experiments. Despite limitations due to the use of phantoms, results show that the proposed segmentation system out-performs the other algorithms tested for comparison.

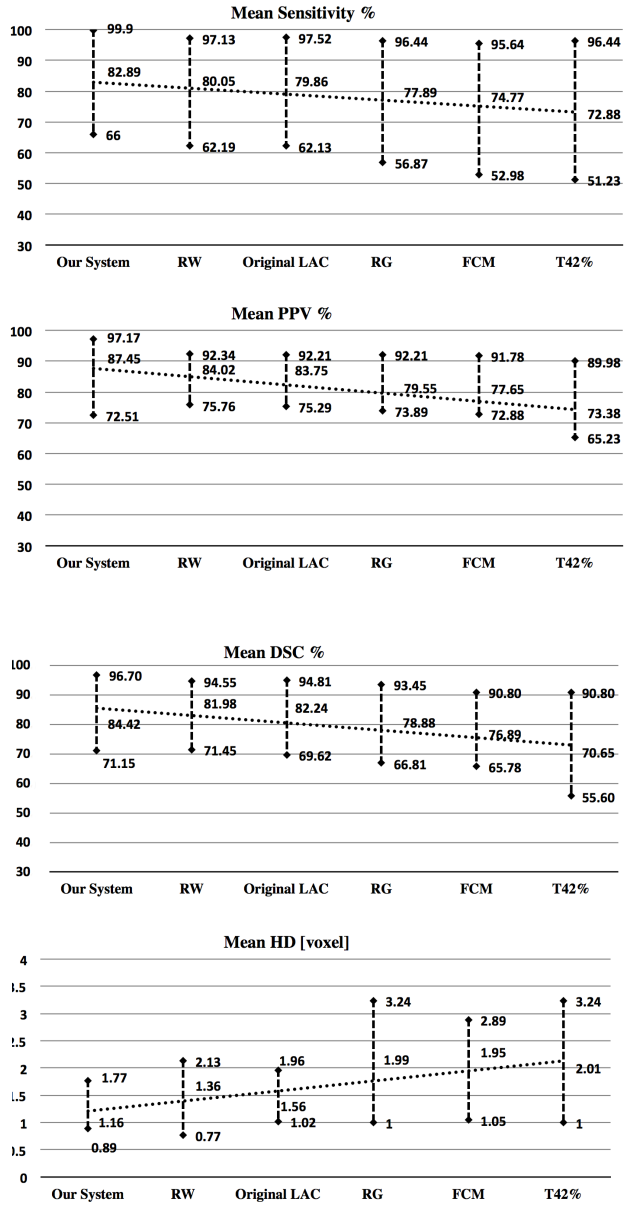


Figure 6.3.2 Sensitivity, PPV, DSC, and HD comparisons (mean and range) of the proposed system to the ones commonly used in the literature in delineation of PET images.

6.3.1.7 Clinical testing and Results on Patients

The evaluation of the segmentation system presented in this study was performed retrospectively (i.e. data were acquired before beginning of the radio-therapy and only at later time employed for the purposes of this study), on 50 tumours: in detail, 10 patients with lung cancer (FDG PET), 25 patients with head & neck cancers (FDG PET), and 15 patients with brain metastases (MET PET).

Performance was evaluated against a "ground truth" manual delineation provided by three expert operators [62] [36]. Automatically segmented BTVs showed high

agreement with the ground truth delineations (the determination coefficient R2 = 0.98, see figure 5 Figure 6.3.3).

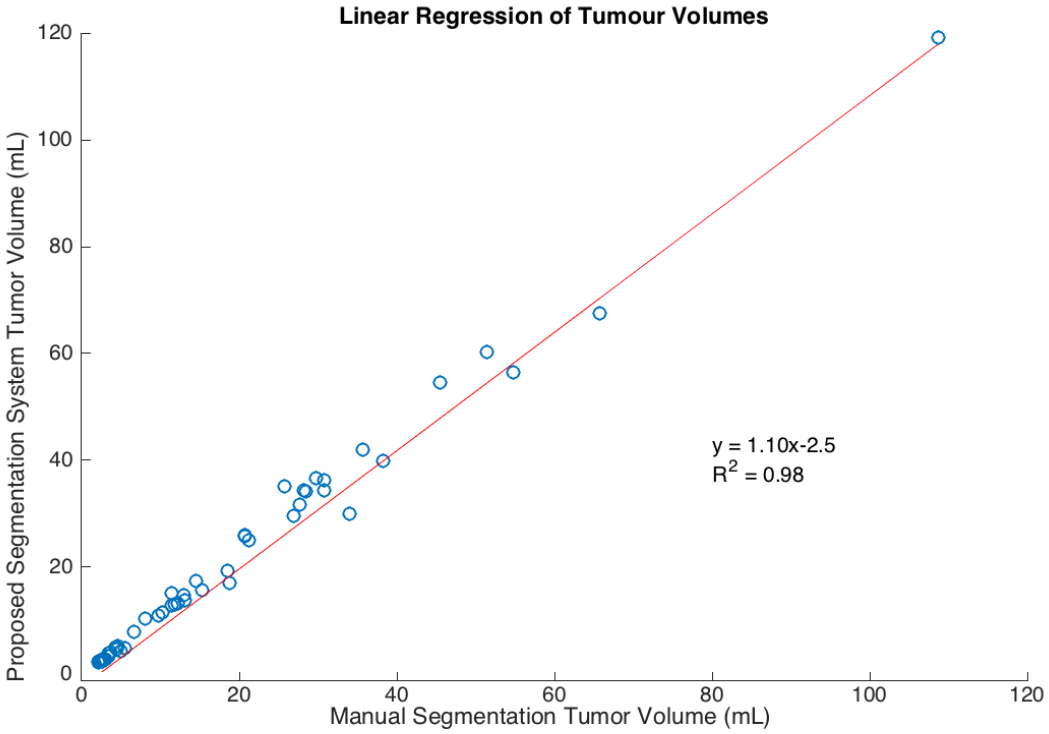


Figure 6.3.3 Linear regression of manually and automatically segmented tumour volumes. The determination coefficient R2=0.98 demonstrates high correlation between them. The first-order coefficient 1.10 indicates that segmentation slightly overestimates tumour volume. The constant offset of −2.5 indicates that automatic segmentation may not identify small lesions (<2.5 ml).

To assess the performance level represented by automatically and manual segmentations provided by three expert operators, we used STAPLE algorithm [36]. Results are provided in Table 6.3.3 as "mean ±standard deviation (±CI) [±CV] of the combination of sensitivity and PPV (see previous section).

Table 6.3.3 Performance level of the three manual segmentations and of the proposed automatic segmentation.

	Score	Score RT	Score FU
Expert 1	95.76%±0.23(±0.27)[±0.25]	95.17%±0.31(±0.35)[±0.33]	96.36%±0.41(±0.46)[±0.42]
Expert 2	87.41%±0.88(±1.00)[±1.01]	84.89%±0.95(±1.07)[±1.12]	89.92%±0.82(±0.93)[±0.91]
Expert 3	89.84%±1.85(±2.09)[±2.06]	87.94%±2.42(±2.74)[±2.71]	91.73%±1.28(±1.45)[±1.40]
Our System	90.68%±1.34(±1.52)[±1.48]	90.80%±1.35(±1.53)[±1.49]	90.57%±1.34(±1.51)[±1.47]

Table 6.3.4, Table 6.3.5 and Table 6.3.6 show sensitivity, PPV, DSC, HD, and MHD results for the three lesion datasets (lung, head & neck, and brain). Since in the Cannizzaro Hospital dataset the segmented BTVs are greater than 2.5 ml (lesions with a sphere-equivalent diameter greater than 17 mm), no distinction between lesion volumes is reported as for phantom sphere results in the previous section. In addition, most segmented BTV (~90%) are homogenous (one single peak is visually detected on the histogram of lesion voxel values). Figure 6.3.4 reports the quantitative comparison between the operator independent segmentation and the gold-standard.

Table 6.3.4 Mean Sensitivities, PPVs, DSCs, HDs, and MHDs for 10 lung cancer studies using FDG-PET examinations are reported. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

Lung Cancer	**Sensitivity [%]**	**PPV [%]**	**DSC [%]**	**HD [voxels]**	**MHD [voxels]**
#1	91.60%	81.70%	86.40%	2.49	1.60
#2	97.60%	82.80%	89.60%	1.98	1.39
#3	98.50%	71.60%	82.90%	2.59	2.09
#4	85.00%	85.00%	85.00%	1.08	0.83
#5	98.90%	83.20%	90.40%	1.48	0.72
#6	88.80%	77.20%	82.60%	1.69	0.84
#7	91.00%	73.60%	81.40%	1.14	0.48
#8	96.70%	78.40%	86.60%	1.25	0.91
#9	94.70%	76.30%	84.50%	2.60	0.67
#10	93.50%	76.60%	84.20%	2.38	0.60
Mean	93.63%	78.64%	85.36%	1.87	1.01
± std	±4.55%	±4.41%	±2.94%	±0.62	±0.51
± CI (95%)	±2.82%	±2.73%	±1.82%	±0.38	±0.32
± CV	±4.86%	±5.60%	±3.44%	±0.33	±0.51

Table 6.3.5 Mean Sensitivities, PPVs, DSCs, HDs, and MHDs for 25 head & neck cancer studies using FDG-PET examinations are reported. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

Neck & head Cancer	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]	MHD [voxels]
#1	98.70%	81.30%	89.10%	1.46	0.43
#2	83.30%	89.70%	86.40%	1.00	0.69
#3	89.80%	76.20%	82.40%	1.07	0.74
#4	85.50%	94.60%	89.80%	0.85	0.59
#5	93.20%	74.80%	83.00%	1.17	0.36
#6	96.20%	81.30%	88.10%	1.42	1.41
#7	93.90%	89.90%	91.80%	1.19	1.18
#8	99.50%	71.80%	83.40%	2.79	0.60
#9	96.90%	78.80%	86.90%	2.46	1.37
#10	97.50%	77.10%	86.10%	2.59	0.72
#11	73.10%	90.50%	80.90%	1.00	0.71
#12	99.60%	73.70%	84.70%	1.40	0.69
#13	93.80%	73.80%	82.60%	1.03	0.94
#14	91.40%	86.50%	88.90%	1.00	0.78
#15	88.80%	74.60%	81.10%	2.78	2.28
#16	77.60%	84.90%	81.10%	1.07	1.01
#17	89.80%	86.30%	88.00%	1.14	0.75
#18	91.90%	87.00%	89.40%	0.90	0.81
#19	91.40%	80.00%	85.30%	1.37	1.27
#20	96.30%	83.90%	89.70%	0.90	1.16
#21	97.30%	80.70%	88.20%	1.88	1.98
#22	80.80%	95.10%	87.40%	0.94	1.07
#23	78.70%	92.50%	85.10%	0.93	1.00
#24	93.30%	70.00%	80.00%	1.33	1.52
#25	96.60%	84.30%	90.00%	0.99	0.58
Mean	91.00%	82.37%	85.98	1.39	0.99
± std	±7.33%	±7.30%	±3.40%	±0.61	±0.46
± CI (95%)	±2.87%	±2.86%	±1.33%	±0.24	±0.18
± CV	±8.05%	±8.86%	±3.95%	±0.44	±0.47

Table 6.3.6 Mean Sensitivities, PPVs, DSCs, HDs, and MHDs for 15 brain cancer studies using MET-PET examinations are reported. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

Brain Cancer	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]	MHD [voxels]
#1	93.50%	89.90%	91.70%	0.62	1.06
#2	90.10%	78.60%	83.90%	1.33	1.14
#3	88.40%	82.80%	85.50%	1.54	0.97
#4	76.75%	99.18%	86.33%	0.58	0.61
#5	86.47%	95.89%	90.83%	0.58	0.56
#6	96.20%	75.60%	84.70%	1.91	1.18
#7	91.80%	79.50%	85.20%	2.37	1.22
#8	94.50%	84.70%	89.40%	0.79	0.85
#9	86.80%	83.30%	85.00%	1.24	1.05
#10	91.20%	88.70%	90.00%	1.40	0.93
#11	91.80%	82.90%	87.10%	1.54	0.71
#12	85.20%	97.00%	90.70%	1.38	0.59
#13	93.20%	90.30%	91.70%	1.54	1.07
#14	93.00%	84.70%	88.70%	1.81	1.72
#15	93.60%	85.80%	89.50%	0.65	0.57
Mean	90.17	86.60%	88.02	1.28	0.95
± std	±4.89%	±6.89%	±2.75%	±0.54	±0.32
± CI (95%)	±2.47%	±3.49%	±1.39%	±0.28	±0.16
± CV	±5.42%	±7.95%	±3.12%	±0.42	±0.33

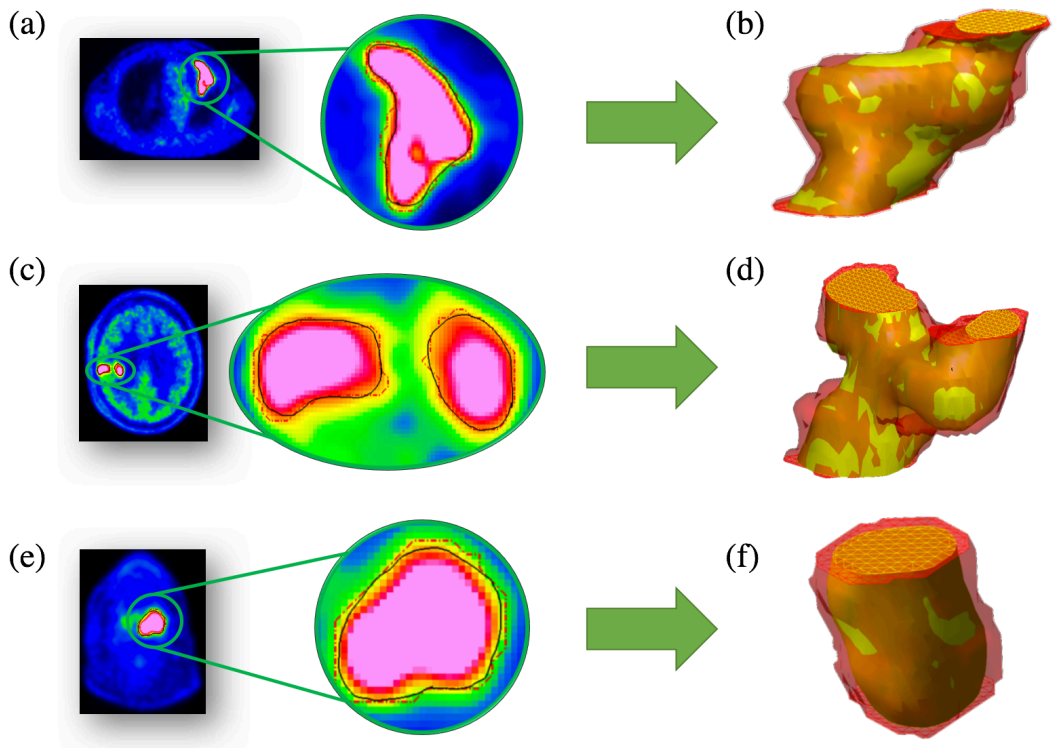


Figure 6.3.4 Three results achieved by the proposed segmentation method on PET images (lung cancer #2, head and neck cancer #25, and brain cancer #13) are reported in figure (a, c, e), respectively. At the bottom right or left, the lesion regions are zoomed. The proposed segmentations (red contours) and the gold standards (black contours) are superimposed. The three-dimensional reconstructions of the tumours are shown in figure (b, d, e): manual (yellow) and proposed (red) BTVs are rendered with transparent surfaces to emphasize volume intersections. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In addition, in Table 6.3.7 the proposed segmentation system is compared with the original LAC and RW methods (both methods obtained similar results and outperformed T42%, RG and FCM methods in phantom experiments, for this reason, T42%, RG and FCM methods were not considered in patient studies). The proposed segmentation yielded better performance in minimizing the difference between manual and automated BTVs than the other two reference methods. As a final test, in order to confirm that the system is indeed user-independent, a sub-dataset of 15 lesions (5 lung cancers, 5 head & neck cancers, and 5 brain metastases) was independently segmented using the proposed system by five different operators (the three experts mentioned above and two non-specialists). The result consisted of five identical BTVs. Without the RG pre-segmentation descripted in the Section 5, BTVs changed less than 6% among users using manual ROIs drawn in the slice with the SUV_{max}^{VOI} by the five users in the above mentioned sub-dataset. Changing more significant (up to 50%) occurred if the users chose different PET slices among them to contour the lesion.

Table 6.3.7 Sensitivities, PPVs, DSCs and HDs for cancer studies using the proposed system, original LAC and RW methods.

Cancer	Sensitivity [Mean ± std]	PPV [Mean ± std]	DSC [Mean ± std]	HD [Mean ± std]
	Our System			
Lung	93.63 ± 4.55%	78.64 ± 4.41%	85.36 ± 2.94%	1.87 ± 0.62
Head & Neck	91.00 ± 7.33%	82.37 ± 7.30%	85.98 ± 3.40%	1.39 ± 0.61
Brain	90.17 ± 4.89%	86.60 ± 6.89%	88.02 ± 2.75%	1.28 ± 0.54
	Original LAC			
Lung	92.68 ± 5.67%	71.92 ± 12.36%	80.33 ± 9.51%	3.29 ± 0.74
Head & Neck	80.32 ± 10.82%	83.88 ± 12.06%	78.73 ± 6.68%	2.81 ± 0.49
Brain	91.16 ± 6.16%	77.38 ± 13.89%	83.55 ± 9.03%	2.88 ± 0.54
	RW			
Lung	92.55 ± 5.66%	74.72 ± 7.44%	82.03 ± 6.52%	2.29 ± 0.77
Head & Neck	79.03 ± 7.29%	84.58 ± 9.57%	82.59 ± 4.95%	1.41 ± 0.40
Brain	88.12 ± 8.36%	86.68 ± 3.43%	86.75 ± 5.14%	1.31 ± 0.64

6.3.1.8 Discussion

PET quantification and segmentation are crucial for the accurate interpretation of clinical PET images and are of foremost importance to obtain operator-independent evaluations and results. The repeatability of the BTV delineation may be ensured only by using computer-assisted methods. For this reason, lesion delineation must be obtained using algorithms able to yield an accurate segmentation with consequent accurate calculation of parameters such as SUV, BTV, and TLG, minimizing operator-dependence and increasing result repeatability.

The key novelty of the proposed approach is the integration of different methodologies that have been appropriately combined and adapted to the PET imaging to obtain a new, smart system for the delineation of an operator independent BTV to be used for RTP or therapy evaluation purposes. The system achieves a high degree of automation without the need for any user-defined parameter.

First, we integrate the body-weighted SUV [102] in our system to pre-process the uptake data of PET images. The SUV normalizes the voxel activity considering acquisition time, administered activity, and patient's weight. Others PET parameters (i.e., SUVlbm or SULpeak [3]) could be used rather than body-weighted SUV as well with minor changes to the system that would still lead a convergent, fully repeatable, and user independent segmentation. We use the body-weighted SUV because it is the most common parameter used in our nuclear medicine department.

Second, while the system still exploits an initial ROI provided by the user to avoid false positive or critical conditions, an automatic pre-segmentation process is used to compute a robust user independent ROI in the vicinity of the user input. This is then automatically fed as input into the detailed segmentation algorithm. Third, segmentation is performed using a slice-by-slice marching approach based on the

LAC method appropriately modified to support PET images. And as final feature a fully automatic stop condition is provided. Due to the high level of automation for the whole process, the final results become completely independent of variation in the initial user input, as confirmed by the inter operator segmentation test in Section 5.

The main reason to require the user to provide an initial ROI is because the distribution of FDG is not limited to malignant tissue. FDG enters the cells according to glucose transport mechanism. For this reason, in order to avoid healthy but active structures where a high radiotracer uptake is to be considered normal, the target lesions must be highlighted by the operator. The region growing is used to "expand" the initial user-provided manual segmentation (performed on one slice) to a three-dimensional volume enclosing the anomalies of interest. We chose region growing because it is able to identify a contour encircling the tumour area starting from a single seed point (the voxel with SUV_{max}^{j}) always obtaining the same starting contour (Figure 6.3.1 e). The use of alternative segmentation methods for the computation of the user independent ROI would lead to a slight BTV variation. However, the identification of the hottest slice (the slice with SUV_{max}^{VOI}) in which to compute the user independent ROI is mandatory to avoid a great variability of the final segmentations (see Section 5). The proposed pre-segmentation process used to identify the hottest slice and, consequently, to compute the user independent ROI is crucial to obtain repeatable results.

Further, in the case of brain lesions, as reported in [84], a fully automatic method can be implemented by taking advantage of the great sensitivity and specificity of MET radiotracers in the discrimination between healthy versus cancer. Hence initial user input, which we require in FDG-PET studies, could be completely avoided in brain studies.

Performance measurements of the proposed segmentation system have been obtained by phantom experiments consisting of hot spheres in a warm background. While such phantoms don't replicate all the properties of real tissue, they nevertheless represent a useful tool to assess common reference performances across different algorithms. One of the main drawbacks of this validation is the plastic wall which separates the spheres from the background. Indeed, PET image quantification should be carried out in phantoms with inactive background to reduce the cold wall effect [88]. Nevertheless, this condition is too different from actual clinical conditions; the model proposed by [88] may not be useful to test segmentation algorithms [138]. To ease the problem, we insert hot spheres in a warm background. Another drawback of the body phantom validation comes from the use of homogeneous spheres. However, in our patient studies, almost all the tumours are characterized by a fairly homogeneous distribution of uptake. A single peak is generally detected on the histogram of segmented lesions (see Section 5). In this context, therefore, segmentation performance using phantom experiments can be considered a suitable tool to judge the reliability of the method.

In particular, the proposed segmentation system showed high degree of similarity (DSC and sensitivity greater than 90% for the spheres with a diameter greater than 17 mm) and provided better results in minimizing the difference between actual and automated BTVs than the other state-of-the-art methods.

Nevertheless, when lesion sizes are smaller than 2–3 times the FWHM of the point spread function of the PET image resolution reconstructed by the PET imaging system, the under-estimation of metabolic activity due to PVE cannot be assumed to be negligible. The separation of small oncological lesions from the background region is still very challenging [112].

Concerning patient studies, knowledge of the ground truth is impossible to obtain while patients are still undergoing radiotherapy, and unfortunately, the only valid ground truth for quantitative segmentation evaluation is obtained by histological investigations, in general performed after that long radio-therapic cycles altered the cancer mass morphology. Although histological specimen is used in some studies [120][121], the approach may be problematic because irregular contractions can occur during tissue fixation. Consequently, manual delineation of three experts was used as a gold-standard. Nevertheless, the manual BTV delineation is a challenging task even for expert physicians in deciding which regions to include or exclude in the RTP. It is different between operators (for example, radiotherapy planning experts tend to draw larger contours than nuclear medicine physicians and that obviously has a strong impact on the resulting surrogate of truth). For this reason, to assess the applicability of the proposed algorithms in a clinical environment and variable conditions (radiotracers and body districts), manual segmentations were used to produce a consolidated reference (as suggested in [36]).

Fifty patients that had been referred for a diagnostic PET/CT scan before radiotherapy treatments have been considered. Patients underwent PET examinations with FDG or MET radiotracers because different body districts have been considered (10 lung cancers, 25 head & neck cancers and 15 brain metastases). Results show that the proposed segmentation approach can be considered clinically feasible, since it has been integrated in the current clinical practice. In addition, automatically segmented tumour volumes showed high agreement with the manual segmentations ($R2 = 0.98$). Considering that the proposed system has been implemented in the Matlab R2016a environment and comfortably runs on a standard iMac machine, a qualified IT staff could integrate the system in the PET workstations, as a built-in PET tool for use in medical environment, with a minimum effort. This solution could significantly improve imaging workflow allowing clinicians to include BTV information into the RTP, so obtaining a therapy customized on the patient. In addition, while in our study the user initial input was provided by mouse, user input experience could be easily enhanced using alternative input methods (e.g. stylus, touch-screen, etc.), without affecting the result.

Finally, the proposed approach is different with respect to joint segmentation approaches (PET-CT-MR). The assumption of a “ground truth” joint volume,

defined on fused multimodal imaging data is sometimes misleading. The assumption that a one-to-one correspondence between metabolic and anatomical regions exists is unrealistic [86]. Lesions may present smaller uptake regions compared to anatomical region. In the same way, the PET lesion may show additional area compared to lesion boundaries in CT or MR images. Since different imaging modalities could convey different information, multimodal segmentation may actually compromise the quality of segmentation producing disagreement between anatomical and metabolic boundaries. As a result, the tumour volumes defined on PET and on CT or MRI could be highly different [135]. For this reason, multimodal studies must be combined in a smart fashion with a customization/adaptation of the segmentation method according to the specific clinical scenario. For example, in the clinical case shown in [114], semi-automatic BTV radically changed the RTP because uptake is found outside the anatomical volume in an involved lymph node which is not visible in CT images. However, the extraction of anatomical structures by CT examination can still convey useful information to locate health tissues (i.e. brain, heart, bladder, and kidneys) so to avoid ambiguities and false positives and removing the confounding contribution of FDG-avid normal tissues from the analysis. In addition, further investigations will be carried out to assess the prognostic usefulness and long-term clinical impact of the prosed system, comparing the extracted BTV with clinical outcomes, progression-free survival and overall survival.

6.3.2 Normal and abnormal tissue classification in positron emission tomography oncological studies

In the current section, firstly PET imaging data concerning the patients are reported, and then our classification system is described.

6.3.2.1 Patient Study

Eighty oncological patients (30 females, 50 males, mean age 56.5 years) have been retrospectively evaluated. In particular, 25 lung tumors, 20 HNC, 15 bone tumors, 15 lymph node metastasis, and 5 bladder tumors have been considered in order to investigate many types of abnormalities.
FDG PET/CT scans are performed in accordance with the standard whole-body oncological protocol in use in our institution, following international guide- lines published on behalf of the European Association of Nuclear Medicine (EANM) [17] [139]. Written informed consent is obtained by each patient before the examination and the study is performed in compliance with the Declaration of Helsinki. FDG PET/CT is per- formed in the fasting state for at least 6 h and the glucose level is always lower than 160 mg/dL. Images are acquired on a PET/CT Discovery scanner (GE Healthcare, Haifa, Israel) from the vertex to the mid- thigh, with inclusion of the

upper extremities. Acquisition is started 60 min after the intravenous administration of 3.7 MBq/Kg FDG (6–8 beds, 2–4 min per bed position). A low-dose CT (90–120 mA, 140 kV, 0.8 s per tube rotation) is also acquired to perform non-uniform attenuation correction. FDG PET/CT images are reconstructed to a 256 × 256 matrix and qualitatively evaluated by an experienced nuclear medicine physician: abnormal tissue is identified if the metabolic activity in the area is moderately or markedly increased relative to normal surrounding tissues. In addition, abnormal and normal tissues are selected on PET images by the physician using the semi-automatic method described in the next section.

6.3.2.2 Results

To test the effectiveness of the proposed system, a direct comparison among six different methods has been performed. In one case, vectors are labeled as normal or abnormal tissues using the semi-automatic method; in other cases, FCM and k-means are used for labelling vectors in an unsupervised manner. The Linear KSVM and KNN classification techniques are applied for each of the data set. The performance of the methods is also evaluated after PCA-based dimensionality reduction.
Performance measures are calculated regarding correct/incorrect classification:

- True Positive (TP): a vector is considered Abnormal both in medical classification and in classification method;
- False Positive (FP): a vector is considered Abnormal in classification method and Normal for medical classification;
- True Negative (TN): a vector is considered Nor- mal both in medical classification and in classification method;
- False Negative (FN): a vector is considered Abnormal in classification method and Normal for medical classification.

At this point it is possible to calculate the Sensitivity, Specificity, Negative Predictive Value, Precision, Accuracy, and Error scores:

- Sensitivity: It is defined as the percentage of effective abnormal tissues that are correctly identified as such:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

- Specificity: It is defined as the percentage of effective normal tissues that are correctly identified as such:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- Negative Predictive Value: It is defined as the probability that vector with a normal test are truly not abnormal:

$$\text{Negative Predictive Value} = \frac{TN}{TN + FN}$$

- Precision: It is defined as, related to reproducibility and repeatability, the degree to which repeated classifications under unchanged conditions show the same results:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Accuracy: It is defined as the degree of closeness of classifications of a abnormal to that abnormal's true medical classification:

$$\text{Accurancy} = \frac{TP + TN}{TP + FP + FN + TN}$$

- Error: It is defined as the inverse of accuracy:

$$\text{Error} = \frac{FP + FN}{TP + FP + FN + TN}$$

The implementation of the proposed method has been running on a general purpose PC with a 2.3 GHz Intel Core i5 processor, 8 GB 1333 MHz DDR3 memory, and Mac OS × 10.8.5 version. An analysis of the time performance show that algorithms are real time: the training and validation time for all methods is less than 1.2 s. The results show that KNN classification obtains excellent results, as depicted in Table 6.3.8.

Table 6.3.8 Classification comparison of the 6 different methods: KSVM and KNN classification techniques are applied for each of the three labeling techniques; the 5-fold cross-validation has been integrated to make reliable the classifier.

Labeling	Semi-automatic %	Fuzzy C-Means	K-Means	Semi-automatic	Fuzzy C-Means	K-Means
Sensitivity Eq. (1)	84.98	97.15	99.02	86.25	100.00	100.00
Specificity Eq. (2)	85.54	68.31	59.71	90.00	63.81	58.10
Negative Predictive Value Eq. (3)	85.07	96.23	98.37	88.37	100.00	100.00
Precision Eq. (4)	85.45	75.75	71.11	89.81	73.64	70.51
Accuracy Eq. (5)	85.26	82.73	79.36	88.12	81.91	79.05
Error Eq. (6)	14.74	17.27	20.64	11.87	18.09	20.95

In particular, for KSVM with semi-automatic Labeling, the whole classification accuracy is 85.26%; for KSVM with Fuzzy C-Means Labeling, is 82.73%; for KSVM with K-Means Labeling, is 79.36%; for KNN with semi-automatic Labeling, the whole classification accuracy is 88.12%; for KNN with Fuzzy C-Means Labeling, is 81.91%; and for KNN with K-Means Labeling, is 79.05%. After PCA-based dimensionality reduction to 23-dimensional vector in order to pre- serve over 95% of total data variance, the performance is poor for all methods (see Table 6.3.9) except for the semi-automatic KNN method where sensitivity, specificity, negative predictive value, precision, accuracy, and error scores are comparable to those without dimensionality reduction. In addition, 5 different training and validation datasets of 130 vectors with unequally distributed classes have been randomly created from the original dataset. In a context without human supervision it is difficult to obtain input data-sets with evenly distributed classes (in our study, 80 lesion and 80 background vectors). The new training datasets have been labeled using k-means and FCM methods. In all cases, both KSVM and KNN performances decrease using these datasets: the sensitivity is greater than 99% (range: 99–100%) and the specificity is less than 53% (range: 38–53%).

Table 6.3.9 Classification comparison of the 6 different methods after PCA-based dimensionality reduction: KSVM and KNN classification techniques are applied for each of the three labeling techniques; the PCA has been used to reduce the vectors dimensionality; the 5-fold cross-validation has been integrated to make reliable the classifier.

	5-Fold Cross Validation					
Classification	PCA + KSVM			PCA + KNN		
Labeling	Semi-automatic	Fuzzy C-Means	K-Means	Semi-automatic	Fuzzy C-Means	K-Means
Sensitivity Eq. (1)	83.96	91.87	91.35	86.12	100.00	100.00
Specificity Eq. (2)	80.21	53.65	61.50	90.31	48.21	52.02
Negative Predictive Value Eq. (3)	83.39	86.82	87.85	87.15	100.00	100.00
Precision Eq. (4)	80.93	66.48	70.45	90.32	65.90	67.59
Accuracy Eq. (5)	82.08	72.76	76.43	88.21	74.10	76.01
Error Eq. (6)	17.92	27.24	23.57	11.78	25.90	23.99

6.3.2.3 Discussion

In this work, we introduced a computerized system for automatic classification of PET tissues using an ad- hoc training set to obtain optimal target and back- ground region identifications. In PET examinations, the automatic identification of lesion boundaries is not a trivial problem. For this reason, the nuclear medicine physician must place fixed ROIs along the lesion boundary so as to separate abnormal from back- ground regions. This approach allows a better classification of lesion tissues by excluding false positives (normal structures like the brain, heart, bladder, and kidneys that normally have high FDG uptake). In our training set, the uptake intensity difference between the lesion and its surrounding is used as key feature to differentiate PET tissues. This is a fundamental, but at the same time, simple and rapid task to obtain an optimal tissue classification. We believe that the combination of clinical knowledge about the lesion boundary localizations with image analysis techniques is extremely useful for the correct classification of PET regions.

Moreover, the output provided by tissue classification can be used as stop/control condition in new or existing segmentation methods to obtain more accurate metabolic volume-based PET parameters (i.e.: SUV mean, TLG, MTV) or more accurate biological target volumes for radio-therapy treatment. The pro- posed system shows real-time and good performance; for this reason the provided output can be used to enhance existing delineation methods in order to obtain most performing results. In particulary, KNN with semi-automatic labeling outperforms the other proposed methods. This is to be expected because the tissue labeling is supervised. However, unsupervised labeling methods achieve good classification results with the benefit of requiring a lower user interaction effort and lower levels of the user's specialist knowledge than the supervised method. Nevertheless, the proposed semi-automatic method is user-friendly and it is not time-consuming. PCA-based dimensionality reduction affects the performance of all methods except for the semi-automatic KNN method. Nevertheless, the reduction of computational load can be advantageous in the clinical environment if the classification output is used as stop/control condition in new or existing segmentation methods although the proposed system without dimensionality reduction shows real-time performance. Finally, 5 different input datasets have been randomly created and labeled by using unsupervised methods to avoid lesion and background evenly distributed classes. In all cases, classification performances decrease: lesion tissues are correctly identified; many background tissues are erroneously identified as lesions. At this moment, supervised tissue labeling is mandatory in PET imaging to obtain good classification scores.

Collection of a large sample of clinical examinations will better determine the adoption of the system in clinical environment; the system can be readily applicable for images produced from different manufacturers, since it worked on standard Digital Imaging and Communications in Medicine (DICOM) format. However,

different PET scanners or different scanning protocols may affect the measured SUVs and, consequently, the final results. In these cases, a parameter normalization would be mandatory. A limitation of this study is the classification of small tissues: partial volume effect (PVE) is the most important factor impacting the quality and the quantitative accuracy in PET studies [77]. The images are blurred due to the limited spatial resolution of PET scanner and small lesions appear larger. Several corrective methods have been developed and the technique described in [140] could be included in order to limit the PVE in our PET studies.

6.3.3 K-nearest neighbor driving active contours to delineate biological tumor volumes

6.3.3.1 Criteria for results evaluation

A framework for the assessment of performance of both the KNN classifier and the proposed segmentation algorithm is presented. The effectiveness of the KNN classification is calculated regarding correct/incorrect classification using sensitivity, specificity, precision, accuracy, and error scores.

The sensitivity is the number of correctly classified positive samples divided by the number of true positive samples:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%$$

where TP and FN are the number of true positives and false negatives, respectively. The specificity is the number of correctly classified negative samples divided by the number of true negative samples:

$$Specificity = \frac{TN}{TN + FP} \times 100\%$$

where TN and FP are the number of true negatives and false positives, respectively.

The precision is related to reproducibility and repeatability and it is defined as the degree to which repeated classifications under unchanged conditions show the same results:

$$PPV = \frac{TP}{TP + FP} \times 100\%$$

The accuracy is defined as the number of correctly classified samples divided by the number of classified samples:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%$$

Finally, the error is defined as the inverse of the accuracy:

$$Error\ = \frac{FP + FN}{TTP + FP + FN + TNP + FP} \times 100\%$$

Concerning the segmentation algorithm performance, according to the formulations presented in [36][100][102], overlap-based and spatial distance-based metrics are considered to determine the accuracy achieved by the proposed method (assessed volume) against the reference volumes (i.e. the gold standard, previous section).

TP, FP, TN, and FN voxels were used to measure the overlap and difference between the two volumes. In particular, mean, and standard deviation of sensitivity, PPV, dice similarity coefficient (DSC), and Hausdorff distance (HD) were calculated.

The true positive volume fraction (TPVF), also called sensitivity (see eq. 13), is the fraction of the total amount of tissue in the proposed segmentation method which overlaps with the reference volume. The ideal segmentation algorithm would be 100 % sensitive (segmenting all voxels from the target voxels) and 100 % specific (not segmenting any from the background voxels). In particular, TPVF [101] is a crucial measure in radiotherapy in order to obtain optimal RTP avoiding cancer recurrence. The true negative volume fraction (TNVF), also called specificity (see eq. 14), is the fraction of tissue defined in the reference volume that is missed by the segmentation system Sseg. Considering that the number of true negatives depends on the space volume I, the specificity has limited use, and only the sensitivity conveys useful information. The specificity can be replaced with the PPV [108], otherwise called precision (eq. 15), which corresponds to the fraction of tissue from the reference volume Sref which overlaps with the segmentation system Sseg. DSC value [100] measures the spatial overlap between the reference volume and the segmentation system:

$$DSC = \frac{2\left|s_{Seg} \cap s_{\mathrm{ref}}\right|}{\left|s_{Seg}\right| + \left|s_{\mathrm{ref}}\right|} = \frac{2{\times}TP}{2{\times}TP + FP + FN} \times 100\%$$

A DSC = 100% indicates a perfect match between two segmentations, while a DSC = 0% indicates no overlap. Nevertheless, overlap-based metrics are highly dependent on the segmentation size. For this reason, distance-based metrics are preferable when the boundary segmentation is critical, such as in BTV delineation for RTP. In particular, HD [36] is used to measure the most mismatched boundary voxels between automatic and manual BTV: a small median of HD means an accurate segmentation, while a large median of HD means no accuracy. The performance of the proposed method is compared to other state of the art BTV segmentation methods. In detail: the fixed thresholding (42% of the SUVmax) (T42%) [95], the RG [110], the FCM clustering [109], the enhanced RW [86], and the original LAC [83]. The proposed segmentation BTVs and evaluation tools were implemented in Matlab® R2016a (MathWorks, Natick, MA, USA), running on an iMac computer with a 3.5 GHz Intel Core i7 CPU, 16 GB 1600 MHz DDR3 memory, and OS X El Capitan.

6.3.3.2 Results KNN validation

As reported in the "Materials and methods" section, the capability of the classifier in discerning different kinds of tissues must be assessed prior of its use in combination with the LAC. Therefore, the K-Fold cross-validation has been integrated in the KNN classifier. The optimal K value has been determined as 5 through the trial-and-error method (k range: 5–15, step size of 5) corresponding to the highest classification accuracy. The validation results show that the KNN classification achieves excellent result with a sensitivity of 97.26%, a specificity of 81.75%, a precision if 98.89%, an accuracy of 95.42%, and an error of 4.58%.

6.3.3.3 Results on Phantoms

As previously introduced, performance of the discussed algorithm in segmenting localized, well defined anomalies is evaluated considering small spheres with diameters of 10, 13, and 17 mm and large spheres with diameters of 22, 28, and 37 mm. Evaluation parameters, such as sensitivity, PPV, DSC, and HD for the two sets of spheres, are summarized in Table 6.3.10 and Table 6.3.11 respectively. Roman numbers are used to label five independent phantom experiments carried out with different signal ratios between sphere and background regions: 1.5:1 for the phantom "I", 3:1 for the phantom "II", 5:1 for the phantom "III", 7:1 for the phantom "IV", and 8:1 for the phantom "V".

Table 6.3.10 Sensitivity, PPV, DSC, HD rates obtained over NEMA IEC phantoms (II–V). Phantoms have the following signal contrast: (II) 3:1, (III) 5:1, (IV) 7:1, and (V) 8:1. Three different spherical lesions of size 10, 13, and 17 mm in diameter are considered to assess the performance of the proposed segmentation method. Mean value and standard deviation (std) are reported in the last rows.

NEMA IEC Phantom	Sphere Diameter	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]
II	10 mm	66,70%	76,20%	71,10%	1,00
III		72,70%	94,10%	82,10%	1,21
IV		84,00%	61,80%	71,20%	0,90
V		78,30%	60,00%	67,90%	1,14
II	13 mm	66,70%	100,00%	80,00%	1,28
III		80,40%	84,10%	82,20%	1,00
IV		73,80%	91,20%	81,60%	0,68
V		70,00%	91,30%	79,20%	1,00
II	17 mm	81,70%	97,10%	88,70%	0,93
III		69,90%	96,30%	81,00%	1,28
IV		74,30%	92,00%	82,20%	1,20
Mean ± std		74,41±5,98%	85,83±13,94%	78,84±6,18%	1,06±0,19

Table 6.3.11 Sensitivity, PPV, DSC, HD rates obtained over NEMA IEC phantoms (I–V). Phantoms have the following signal contrast: (I) 1.5:1, (II) 3:1, (III) 5:1, (IV) 7:1, and (V) 8:1. Three different spherical lesions of size 22, 26, and 37 mm in diameter are considered to assess the performance of the proposed segmentation method. Mean value, and standard deviation (std) are reported in the last rows.

NEMA IEC Phantom	Sphere Diameter	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]
I	22 mm	91,60%	75,53%	82,68%	1,66
II		91,60%	87,50%	89,50%	1,17
IV		94,80%	84,70%	89,50%	1,06
V		96,30%	87,90%	91,90%	1,00
I	26 mm	86,03%	89,67%	87,58%	1,44
III		91,80%	88,80%	90,30%	1,10
IV		94,70%	90,20%	92,40%	1,00
V		93,50%	88,10%	90,80%	1,18
I	37 mm	92,38%	90,01%	90,82%	1,41
II		91,00%	93,00%	92,00%	1,25
III		93,80%	90,20%	92,00%	1,23
IV		93,40%	92,50%	93,00%	1,00
V		92,60%	91,80%	92,20%	1,13
Mean ± std		92,58±2,49%	88,45±4,48%	90,36±2,75%	1,20±0,20

Table 6.3.10 shows the volumetric accuracy for the smaller spheres, while Table 6.3.11 shows the volumetric accuracy for the larger spheres. In Table 6.3.10, the results for the phantom "I" are not reported because the ratio between sphere and background concentrations is too low (the target radioactivity concentration is only more than one-and-a-half times the background radioactivity concentration), and no PET region with high uptake concentration is visible for spheres with diameter < 22 mm. No significant difference between background and target is observed. For the same reason, the accuracy improved for all spheres, regardless of their volume, when the ratio between sphere concentration and background radioactivity concentration was increased. In addition, three phantom spheres, with ratio between sphere and background radioactivity concentrations of 3:1 (diameter = 28 mm), 5:1 (diameter = 22 mm), and 8:1 (diameter =17 mm) are not reported because they were used to train the KNN classifier

In general, due to the partial volume effect, the separation of small lesions from the background region is very challenging, and the difficulty increases in critical conditions, such as cases of low signal contrast. The volumes of smaller spheres are underestimated (mean difference between segmented and actual volumes = -18.83±9.64%) with a prevalence of false negatives when compared to false positives. As expected, large errors occur in lesions less than 2 cm in diameter [77].

Table 6.3.11 shows the results for all spheres with a diameter greater than 17 mm. In all conditions, excluding the phantom 'I' for the aforementioned reason, a DSC rate above 90% and a sensitivity rate greater than 90% are observed. These performances are generally accepted as excellent. A slight oversizing is observed (mean difference between segmented and actual volumes = 6.76±5.77%) nevertheless, larger margins can help to prevent the extension of cancer infiltration in radiotherapy treatments [99].

In addition, despite the limitations posed by the use of phantoms, our results show that the proposed segmentation method outperforms the algorithms used for comparison (RW, original LAC, RG, FCM, and T42%). Figure 6.3.5 reports the quantitative comparison between the semi-automatic segmentation and the gold-standard considering all phantom experiments.

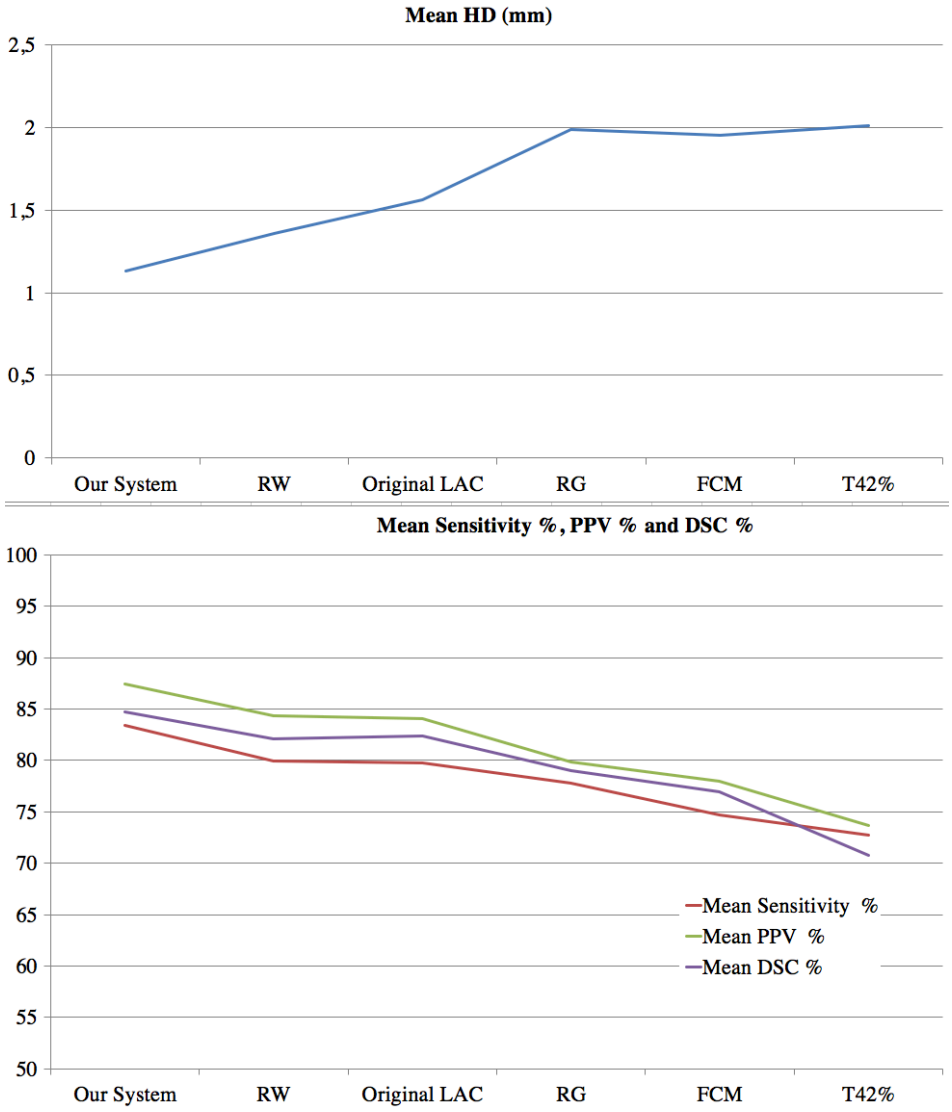


Figure 6.3.5 HD, sensitivity, PPV, and DSC comparisons of the proposed algorithm to the ones commonly used in the literature in delineation of BTVs.

6.3.3.4 Clinical testing and Results on Patients

From the initial set of 58 tumors, eight cases were used to train the KNN. Consequently, the evaluation of the algorithm's performance, obtained by comparing the produced segmentations against the gold standard, involved 50 tumors. In detail, 15 patients with brain metastases (MET PET), 25 patients with HNC (FDG PET), and 10 patients with lung cancer (FDG PET). Table 6.3.12, Table 6.3.13 and Table 6.3.14 show sensitivity, PPV, DSC and HD results for brain metastases, HNC, and lung cancers. Differently from the phantom study, no discussion of the lesion volumes is provided, because all considered BTVs are greater than 2.5 ml (lesions with a sphere-equivalent diameter > 17 mm). In addition, most segmented BTV (~90%) are homogenous (one single peek is visually detected on the histogram of lesion voxel values). Figure 6.3.6 reports the comparison between the operator independent segmentation and the gold-standard. Automatically segmented BTVs showed high agreement with the manually segmented BTVs (the determination coefficient R2 = 0.98).

Table 6.3.12 Mean sensitivities, PPVs, DSCs and HDs for 15 brain cancer studies using MET-PET examinations are reported. Mean value, and standard deviation (std), are reported in the last rows.

Brain Cancer	**Sensitivity [%]**	**PPV [%]**	**DSC [%]**	**HD [voxels]**
#1	93,67%	91,85%	92,60%	0,60
#2	93,00%	81,50%	86,90%	1,06
#3	90,10%	85,40%	87,70%	1,03
#4	95,30%	84,30%	89,28%	0,47
#5	92,71%	88,80%	90,50%	0,53
#6	92,80%	83,70%	88,00%	1,17
#7	91,00%	80,80%	85,60%	2,40
#8	93,80%	88,60%	91,10%	0,50
#9	92,00%	80,00%	85,60%	1,19
#10	93,40%	87,30%	90,30%	1,14
#11	90,10%	86,80%	88,40%	1,51
#12	93,30%	88,50%	90,80%	1,31
#13	93,40%	90,80%	92,10%	1,63
#14	90,40%	87,30%	88,80%	1,08
#15	90,00%	90,80%	90,40%	0,55
Mean± std	92,33%±1,64%	86,43%±3,72%	89,21%±2,17%	1,08±0,52

Table 6.3.13 Mean sensitivities, PPVs, DSCs and HDs for 25 HNC studies using FDG-PET examinations are reported. Mean value, and standard deviation (std) are reported in the last rows.

HNC	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]
#1	94,10%	88,30%	91,10%	1,29
#2	76,20%	91,40%	83,10%	1,00
#3	89,80%	81,60%	85,50%	1,07
#4	85,50%	94,60%	89,80%	0,85
#5	91,50%	83,70%	87,40%	1,00
#6	94,70%	83,30%	88,60%	1,33
#7	88,60%	94,40%	91,40%	1,17
#8	96,60%	81,20%	88,20%	1,54
#9	95,20%	83,30%	88,80%	1,70
#10	88,50%	88,40%	88,40%	1,48
#11	73,10%	90,50%	80,90%	1,00
#12	95,20%	84,00%	89,20%	1,24
#13	91,40%	75,50%	82,70%	1,28
#14	91,40%	86,50%	88,90%	1,00
#15	93,50%	75,90%	83,80%	1,61
#16	77,60%	84,90%	81,10%	1,07
#17	89,80%	86,30%	88,00%	1,14
#18	91,90%	87,00%	89,40%	0,90
#19	82,70%	89,50%	86,00%	1,18
#20	92,60%	89,30%	90,90%	0,57
#21	90,50%	89,20%	89,80%	1,84
#22	89,00%	87,30%	88,10%	0,91
#23	78,70%	92,50%	85,10%	0,93
#24	86,70%	76,50%	81,30%	1,00
#25	89,70%	92,40%	91,00%	0,99
Mean± std	88,58%±6,33%	86,30%±5,37%	87,14%±3,31%	1,16±0,29

Table 6.3.14 Mean Sensitivities, PPVs, DSCs and HDs for 10 lung cancer studies using FDG-PET examinations are reported. Mean value, and standard deviation (std), are reported in the last rows.

Lung Cancer	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]
#1	85,70%	94,10%	89,70%	2,61
#2	90,60%	91,10%	90,80%	2,58
#3	89,90%	89,50%	89,70%	1,34
#4	85,00%	85,00%	85,00%	1,08
#5	95,90%	90,40%	93,10%	1,10
#6	80,80%	82,70%	81,70%	1,54
#7	87,60%	77,20%	82,10%	1,07
#8	91,10%	92,00%	91,50%	0,94
#9	85,30%	90,90%	88,00%	2,16
#10	89,00%	85,50%	87,20%	2,83
Mean ± std	88,09±4,18%	87,84±5,15%	87,88±3,89%	1,73±0,74

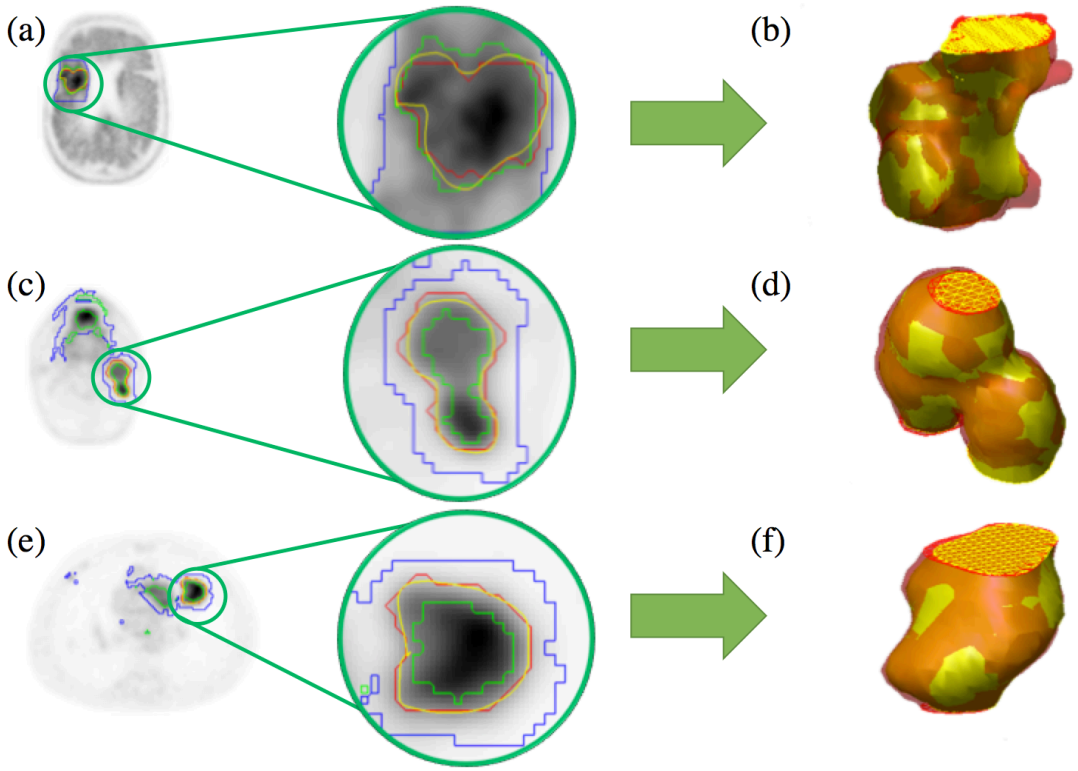


Figure 6.3.6 Three selected segmentation results obtained with the proposed method on PET images (brain cancer #13, head and neck cancer #25, and lung cancer #2) are reported in figure (a, c, e), respectively. For clarity purposes, the lesion region is magnified. The proposed segmentations (red contours) and the gold standards (yellow contours) are superimposed. Blue and green contours concern the KNN tissue classification. The region outside the blue boundary represent the "background", while the region inside the green boundaries, and between the curves, represent the "lesion", and the "border-line" region, respectively. The corresponding three-dimensional reconstructions of the tumors are shown in figures (b, d, e). The manual (yellow), and the proposed BTV (red), are rendered with transparent surfaces in order to emphasize volume intersections.

Table 6.3.15 shows the comparison between the proposed segmentation method and the original LAC and RW methods. Since LAC and RW outperformed T42%, RG, and FCM methods on the phantoms, we decided not to consider the latter group of algorithms in our patient studies. The proposed segmentation method performed better than LAC and RW methods minimizing the difference between manual and automated BTVs. Finally, in order to confirm that the proposed method is operator-independent, a sub-dataset of 5 brain metastases, 5 HNC, and 5 lung cancers was segmented by five different operators (the three experts mentioned above and two non-specialists). Each initial independent manual segmentation led identical BTVs.

Table 6.3.15 Sensitivities, PPVs, DSCs and HDs for cancer studies using the proposed method, original LAC and RW methods.

Cancer	Sensitivity [Mean ± std]	PPV [Mean ± std]	DSC [Mean ± std]	HD [Mean ± std]
	The proposed method			
Lung	88.09 ± 4.18%	87.84 ± 5.15%	87.88 ± 3.89%	1.73 ± 0.74
Head & Neck	88.58 ± 6.33 %	86.30 ± 5.37%	87.14 ± 3.31 %	1.16 ± 0.29
Brain	92.33 ± 1.64%	86.43 ± 3.72%	89.21 ± 2.17%	1.08 ± 0.52
	Original LAC			
Lung	92.68 ± 5.67%	71.92 ± 12.36%	80.33 ± 9.51%	3.29 ± 0.74
Head & Neck	80.32 ± 10.82 %	83.88 ± 12.06 %	78.73 ± 6.68 %	2.81 ± 0.49
Brain	91.16 ± 6.16%	77.38 ± 13.89%	83.55 ± 9.03%	2.88 ± 0.54
	RW			
Lung	92.55 ± 5.66%	74.72 ± 7.44%	82.03 ± 6.52%	2.29 ± 0.77
Head & Neck	79.03 ± 7.29%	84.58 ± 9.57 %	82.59 ± 4.95%	1.41 ± 0.40
Brain	88.12 ± 8.36%	86.68 ± 3.43%	86.75 ± 5.14%	1.31 ± 0.64

6.3.4 Active contour algorithm with discriminant analysis for delineating tumors in positron emission tomography

The dataset used to assess the proposed method comprised a total of six phantom experiments and fifty-eight oncological patients. Some data were used to train the DA classifier, the remaining data to assess the performances of the segmentation algorithm.

6.3.4.1 Phantom Studies

National Electrical Manufacturers Association International Electrotechnical Commission (NEMA IEC) phantom is used for preliminary performance assessment [86][27][87]. The phantom is composed of an elliptical cylinder (D1 = 24 cm, D2 = 30 cm, h = 21 cm) with six different spherical lesions of size 10, 13, 17, 22, 28, and 37 mm in diameter placed at 5.5 cm from the center of the phantom. Both body phantom and spheres are filled with a known amount of radioactive tracer to simulate oncological lesions. The ratio between sphere radioactivity concentrations and background radioactivity concentrations ranged from 1.5:1 to 8:1 for five independent experiments: 1.5:1 for the phantom “I”, 3:1 for the phantom “II”, 5:1 for the phantom “III”, 7:1 for the phantom “IV”, and 8:1 for the phantom “V”. Performance results are calculated considering small spheres (sphere diameter less than or equal to 17 mm) and large spheres (sphere diameter greater than 17mm). This choice was motivated by the fact that large biases are introduced [77] when the lesion size is smaller than 2–3 times the Full Width at Half Maximum (FWHM) of the PET point spread function.

It could be argued that cancer is often heterogeneous and irregularly shaped in contrast with spherical targets and that tracer-filled spheres suffer from cold-body effects when compared to the patient study [78][88]. For this reason, PET phantom acquisition performed by Soffientini et al [89] was included in our experiments. This acquisition has been obtained using eight zeolites (microporous, aluminum-silicate minerals commonly used as commercial adsorbents, which can absorb PET radio-tracer) that allow to generate tumor-like objects with any desired shape, size, and contrast levels without cold walls.
The PET/CT acquisition protocols are described in the in subsequent sections.

6.3.4.2 Clinical studies

In the present study, fifty-eight patients acquired at the Nuclear Medicine Department of the Cannizzaro Hospital (Catania, Italy) have been retrospectively considered. The dataset contains PET scans of seventeen patients with brain metastases, twenty-nine patients with head and neck cancer (HNC), and twelve patients with lung cancers that had been referred to a diagnostic PET/CT scan before radiotherapy treatments.
The institutional Hospital medical ethics review board approved this study protocol and all subjects involved were properly informed and released their written consent. However, segmentations were performed off-line and the results had no influence on the treatment protocol, nor on the patient management. No sensitive patient information was accessed. In FDG PET studies, patients fasted for 12 h before the PET examination, and successively were intravenously injected with FDG. The PET/CT oncological protocol began 60 min after the injection. Patients breathed normally during the PET and CT examinations, and scanning was executed from the top of the skull to the middle of the thigh with the arms along the body. In MET PET studies, for only brain acquisitions, patients fasted for 4 h before the PET examination. The PET/CT protocol began 10 min after the MET injection. The PET/CT acquisition protocol is described in the following section.

6.3.4.3 PET/CT Acquisition Protocol

The acquisitions for both phantom experiments with spheres and clinical studies were performed within the same Nuclear Medicine Department using the same equipment, a Discovery 690 PET/CT scanner with time-of-flight (General Electric Medical Systems, Milwaukee, WI, USA). Time-of-flight imaging method is an innovative and crucial feature in PET scanners to provide a gain in the signal to noise ratio, improving both lesion detectability and uptake measurement. The phantom and

patient protocols included a SCOUT scan at 40 mA, a CT scan at 140 keV and 150 mA (10 sec), and 3D PET scans (2.5 minutes per bed position). PET images were reconstructed using a 3D ordered subset expectation maximization (OSEM) algorithm. All imaging data were encoded in the 16-bit DICOM format. Each PET slice consists of 256×256 voxels with a grid spacing of 2.73 mm3 and thickness of 3.27 mm3. Consequently, the size of each voxel is 2.73 × 2.73 × 3.27 mm3. Thanks to the injected PET radiotracer (FDG or MET), tumors appears as hyper-intense regions. The non-diagnostic CT scan is performed for attenuation correction and anatomic localization of the tumor contextually to PET image acquisition. The CT slice consists of 512 × 512 voxels with size 1.36 × 1.36 × 3.75 mm3.
Zeolite phantom experiment [89] was acquired with Biograph TruePoint 64 PET/CT scanner (Siemens). Attenuation-Weighted-OSEM (AWOSEM) was performed for PET image reconstruction. Four iterations, eight projection subsets, and a 3D Gaussian post-reconstruction filter with 4-mm FWHM were used. Each PET slice consists of 336×336 voxels with size 2.04 x 2.04 x 2.00 mm3.

6.3.4.4 Framework for performance evaluation

A framework for the assessment both in terms of the DA classification and the proposed segmentation algorithm is presented. The performance of the DA is calculated with respect to ratio between successful vs total classifications using sensitivity, specificity, precision, accuracy, and error scores. The sensitivity is the number of correctly classified positive samples divided by the number of true positive while the specificity is the number of correctly classified negative samples divided by the number of true negative samples. The precision is related to reproducibility and repeatability and it is defined as the degree to which repeated classifications under unchanged conditions show the same results:

$$PPV = \frac{TP}{TP + FP} \times 100\%$$

where TP and FP are the number of true positives and false positives, respectively.
The accuracy is defined as the number of correctly classified samples divided by the number of classified samples:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%$$

where TN and FN are the number of true negatives and false negatives, respectively.
Finally, the error is defined as:

$$Error = \frac{FP + FN}{TTP + FP + FN + TNP + FP} \times 100\%$$

About segmentation algorithm performance, according to the formulations presented in [88][89][90], overlap-based and spatial distance-based metrics are considered to determine the accuracy achieved by the proposed method (assessed volume) against the reference volumes (see the next section).

TP, FP, TN, and FN voxels were used to measure the overlap and difference between the two volumes. In particular, mean, standard deviation, 95% confidence interval (CI) and coefficient of variation (CV) of sensitivity, positive predictive value (PPV), dice similarity coefficient (DSC), and Hausdorff distance (HD) were calculated.

The true positive volume fraction (TPVF), also called sensitivity, is the fraction of the total amount of segmented tissue which overlaps with the reference volume s_{ref}. A perfect segmentation algorithm would be 100% sensitive (segmenting all voxels from the target voxels) and 100% specific (not segmenting any from the background voxels). In particular, TPVF [97] is a crucial measure in radiotherapy in order to obtain optimal RTP avoiding cancer recurrence.

The true negative volume fraction (TNVF), also called specificity, is the fraction of tissue defined in the reference volume that is missed by the segmentation system s_{Seg}.

Considering that the number of true negatives depends on the space volume I, the specificity makes little sense and only the sensitivity conveys useful information. The specificity can be replaced with the positive predictive value (PPV) [98].

PPV, also called precision (see eq. 13), is the fraction of the total amount of tissue in the reference volume s_{ref} which overlaps with the segmentation system s_{Seg}.

DSC value [99] measures the spatial overlap between the reference volume s_{ref} and the segmentation system s_{Seg}: a DSC value equal to 100% indicates a perfect match between two volumetric segmentations, while a DSC whose value is 0% indicates no overlap:

$$DSC = \frac{2|s_{Seg} \cap s_{ref}|}{|s_{Seg}| + |s_{ref}|} = \frac{2 \times TP}{2 \times TP + FP + FN} \times 100\%$$

Nevertheless, overlap-based metrics are highly dependent on the segmentation size. For this reason, distance-based metrics are highly recommended when the boundary segmentation is critical, such as in BTV delineation for RTP. In particular, HD is used to measure the most mismatched boundary voxels between automatic and manual BTV: a small median of HD means an accurate segmentation, while a large median of HD means no accuracy [100].

The performance of the proposed method is compared to other state of the art BTV segmentation methods. In particular, the fixed thresholding (42% of the SUVmax) (T42%) [94], the RG [95], the FCM clustering [101], the enhanced RW [86], and the

original LAC [83] methods have been used for comparison. The tool used to provide proposed segmentation BTVs and evaluation task has been implemented in the Matlab R2016a simulation environment (MathWorks, Natick, MA, USA), running on an iMac computer with a 3.5 GHz Intel Core i7 processor, 16 GB 1600 MHz DDR3 memory, and OS X El Capitan.

6.3.4.5 Gold Standard

The match between PET and CT images has been used to evaluate the performance of the proposed method. In patient studies, the histopathology analysis provides the only valid PET ground truth for quantitative assessment [98]. Unfortunately, such information is unavailable after radiotherapy because the treatment destroys the initial geometry. Consequently, the actual gold-standard is impossible to retrieve. For this reason, manual delineation is a commonly accepted substitute for ground truth [78] although it is a critical task. It carries a certain amount of subjectivity, and is often influenced by the clinical specialization of the operator. For example, oncologists will, on average, draw smaller boundaries than radio-therapists. In our study, the chief nuclear medicine physician (M.I. author), the chief radiotherapy physician (M.S. author) and an expert radiotherapy physician (G.R. author) performed manual segmentations in the whole PET dataset. A simultaneous ground truth estimation tool was employed [36], and the segmentations from the three experts were combined to define our "ground truths". From the initial fifty-eight tumor segmentations, eight segmentations were used in the DA training/validation task and fifty ones were used as the "ground truth" for the performance evaluation task of the proposed method. In the same way, three phantom spheres and three zeolites were used in the classifier training/validation task, the remaining ones were used to assess the performance of the present method. In zeolite phantom experiments, gold standards were provided by the authors of the acquisition [89].

6.3.4.6 Results Discriminant analysis validation

The optimal K value of the K-Fold cross-validation integrated in the classifier has been determined through the trial-and-error method as 5 (k range: 5–15, step size of 5). It corresponds to the highest classification accuracy. The results show that the DA classification with both unbalanced and balanced training patient datasets obtains excellent results. The classification results are comparable because the P-value (P-value = 0.99 > Alpha=0.05) leads us to accept the null hypothesis, according to which the averages are equal, as shown in the Table 6.3.16.

Table 6.3.16 Sensitivity, specificity, precision, and accuracy values obtained for the DA validation using both balanced and unbalanced training patient datasets.

Datasets	Sensitivity	Specificity	Precision	Accuracy
Balanced	90.75%	86.70%	99.44%	89.44%
Unbalanced	90.18%	85.94%	99.08%	88.49%

Figure 6.3.7 Shows the ROC analysis performed to measure the DA classification performance after the training step with unbalanced datasets.

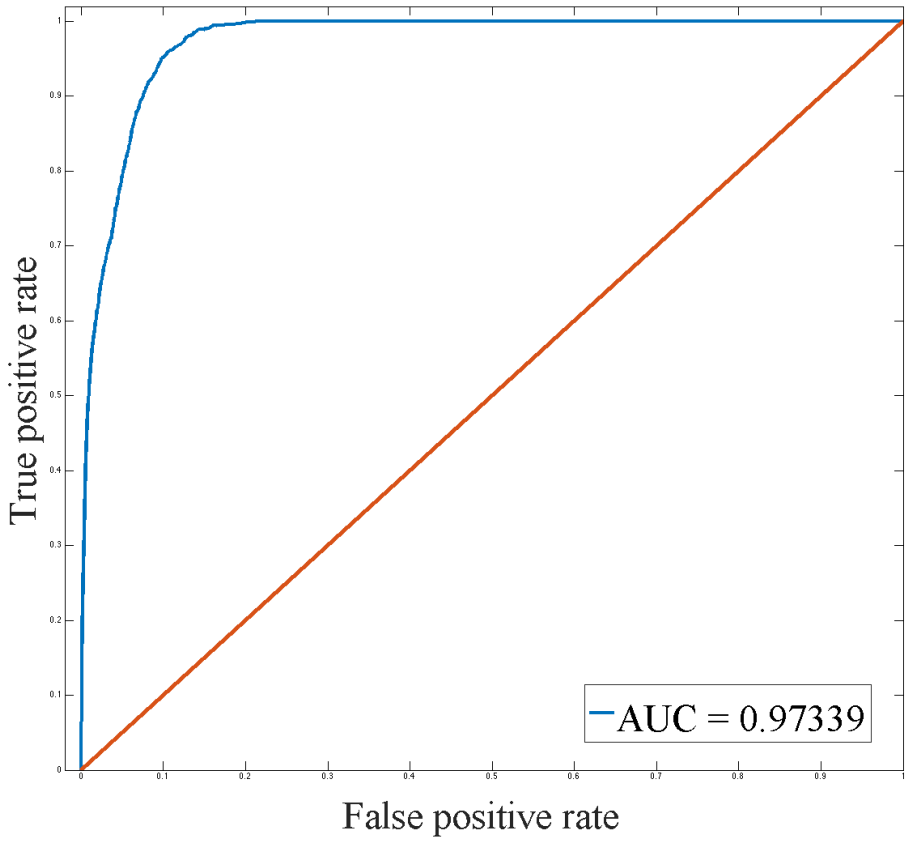


Figure 6.3.7 The classification performance of the DA classifier has been assessed using ROC analysis by calculating the area under the ROC curve (AUC).

6.3.4.7 Clinical testing and Results on Phantoms

Phantom performance results have been divided considering small spheres with diameters less than 22 mm and large spheres with diameter greater than 17 mm (see previous sections). Five independent phantom experiments have been carried out with different signal ratios between sphere and background regions: 1.5:1 for the phantom "I", 3:1 for the phantom "II", 5:1 for the phantom "III", 7:1 for the phantom "IV", and 8:1 for the phantom "V". Table 6.3.17 shows the volumetric accuracy results for the smaller spheres.

Table 6.3.17 Sensitivity, PPV, DSC, HD values obtained over phantom experiments (II–V). Phantoms have the following signal contrast: (II) 3:1, (III) 5:1, (IV) 7:1, and (V) 8:1. Three different spherical lesions of size 10, 13, and 17 mm in diameter are considered to assess the performance of the proposed segmentation method. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

NEMA IEC Phantom	**Sphere Diameter**	**Sensitivity [%]**	**PPV [%]**	**DSC [%]**	**HD [voxels]**
II	10 mm	70.80%	77.30%	73.90%	1.00
III		77.30%	94.40%	85.00%	1.00
IV		75.80%	73.50%	74.60%	1.00
V		78.30%	60.00%	67.90%	1.14
II	13 mm	76.77%	98.48%	86.28%	1.08
III		80.40%	84.10%	82.20%	1.00
IV		73.80%	91.20%	81.60%	0.68
V		70.00%	91.30%	79.20%	1.00
II	17 mm	81.70%	97.10%	88.70%	0.93
III		69.90%	96.30%	81.00%	1.28
IV		75.20%	92.10%	82.80%	1.20
Mean		75.45%	86.89%	80.29%	1.03
± std		±4.02%	±12.04%	±6.07%	±0.16
± CI (95%)		±2.38%	±7.12%	±3.59%	±0.09
± CV		±5.33%	±13.86%	±7.56%	±0.15

The results for the phantom "I" are missing because the contrast between sphere and background is too low (the target radioactivity concentration is only more than one-and-a-half times the background radioactivity concentration). Consequently, no PET regions with high uptake concentration are visible for the smaller spheres. For the same reason, the performance improved for all spheres, regardless of their volume, when the ratio between sphere concentration and background radioactivity concentration was increased. Further three phantom spheres are missing because they have been used to train the DA classifier (see section 5). For this reason, the spheres with ratio between sphere and background radioactivity concentrations of 3:1 (diameter = 28 mm), 5:1 (diameter = 22 mm), and 8:1 (diameter =17 mm) have been excluded.

In general, due to the partial volume effect (PVE), the separation of small lesions from the background region is very challenging, and the difficulty increases in critical conditions (i.e. low signal contrast). The mean difference between segmented and actual volumes is -17.66±7.84%. Volumes of smaller spheres are underestimated with more false negatives than false positives. As expected, large errors occur in the smaller spheres [77].

Table 6.3.18 shows the results for the larger spheres. In all conditions, excluding the phantom 'I' for the aforementioned reason, a DSC rate above 90% and a sensitivity rate greater than 90% are observed. These results are generally accepted as excellent.

Table 6.3.18 Sensitivity, PPV, DSC, HD values obtained over phantom experiments (I–V). Phantoms have the following signal contrast: (I) 1.5:1, (II) 3:1, (III) 5:1, (IV) 7:1, and (V) 8:1. Three different spherical lesions of size 22, 26, and 37 mm in diameter are considered to assess the performance of the proposed segmentation method. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

NEMA IEC Phantom	**Sphere Diameter**	**Sensitivity [%]**	**PPV [%]**	**DSC [%]**	**HD [voxels]**
I	22 mm	88.63%	75.70%	81.65%	1.83
II		91.20%	88.00%	89.60%	1.17
IV		94.80%	84.70%	89.50%	1.06
V		96.30%	87.90%	91.90%	1.00
I	26 mm	88.90%	81.75%	85.18%	1.83
III		91.80%	88.80%	90.30%	1.10
IV		94.70%	90.20%	92.40%	1.00
V		93.50%	88.10%	90.80%	1.18
I	37 mm	89.01%	90.96%	89.97%	1.60
II		90.90%	93.10%	92.00%	1.25
III		93.80%	90.20%	92.00%	1.23
IV		93.40%	92.50%	93.00%	1.00
V		92.60%	91.80%	92.20%	1.13
Mean		92.27%	87.98%	90.04%	1.26
± std		±2.45%	±4.84%	±3.23%	±0.30
± CI (95%)		±1.33%	±2.63%	±1.76%	±0.16
± CV		±2.66%	±5.50%	±3.59%	±0.24

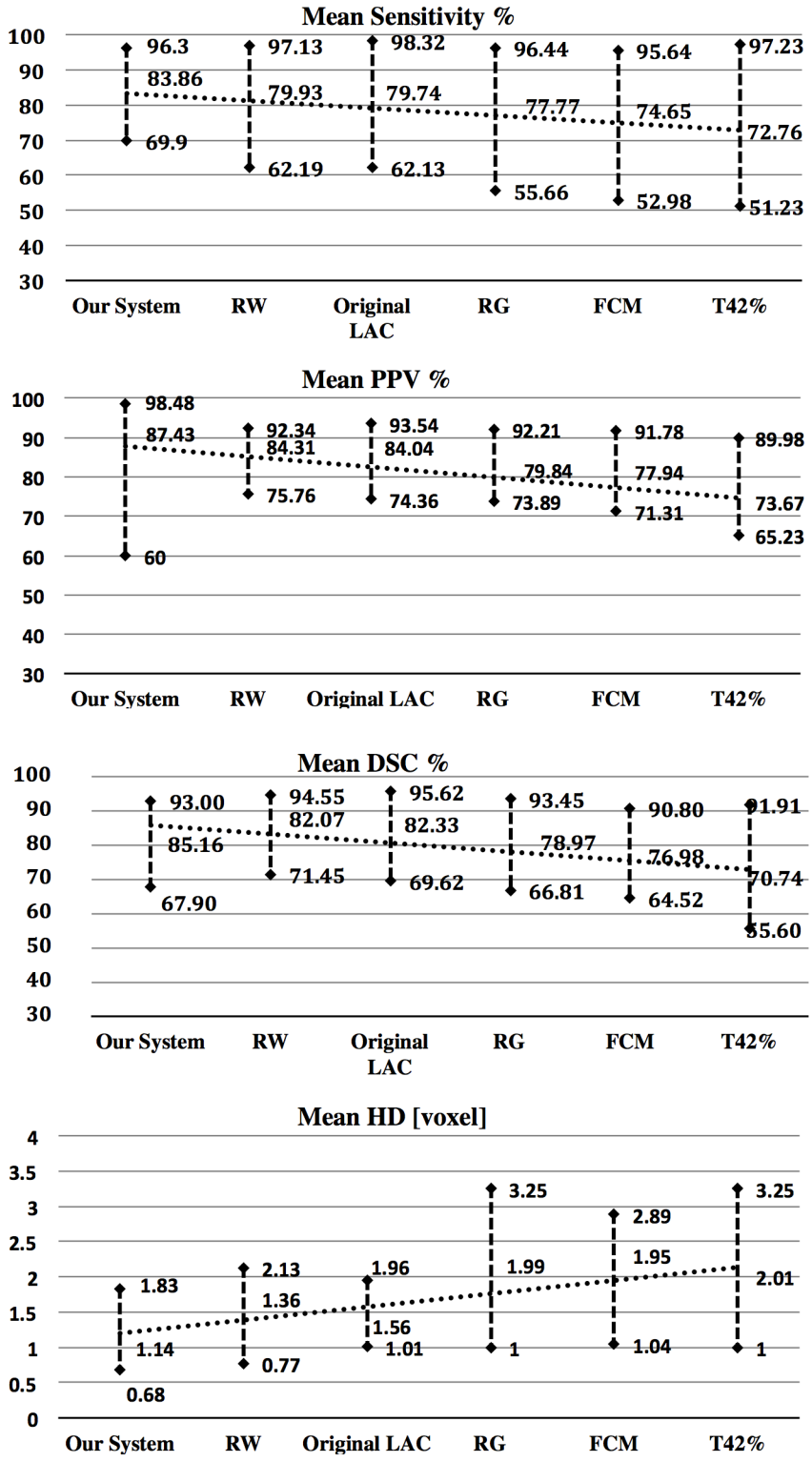


Figure 6.3.8 Sensitivity, PPV, DSC, and HD comparisons (mean and range) of the proposed algorithm to the ones commonly used in the literature in delineation of BTVs.

The mean difference between segmented and actual volumes is 6.34±5.21%. Nevertheless, a slight oversizing can help to prevent the extension of tumor infiltration [110].
In addition, results show that the proposed algorithm outperforms the other ones tested for comparison (RW, original LAC, RG, FCM, and T42%), despite the limitations due to the use of phantoms. Figure 6.3.8 reports the quantitative comparison between semi-automatic segmentations and gold-standards considering all phantom experiments.
In addition, the zeolite phantom acquisition performed by [89] was used to validate our method using heterogeneous and irregularly shaped inserts. Three zeolites have been used to train the DA classifier (a heterogeneous zeolite -#2- and two homogeneous zeolites -#1 and #6 [89]). Results for the remaining five cases are shown in the Table 6.3.19.

Table 6.3.19 Sensitivity, PPV, DSC, HD values obtained over zeolite phantom experiment. Five different zeolites are considered to assess the performance of the proposed segmentation method: #4 and #5 heterogeneous zeolites and #3, #7, and #8 homogeneous zeolites [89]. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

Zeolite ID [25]	**Sensitivity [%]**	**PPV [%]**	**DSC [%]**	**HD [voxels]**
#3	66.28%	97.98%	79.07%	0.35
#4	85.66%	73.67%	79.21%	0.57
#5	82.35%	66.40%	73.52%	0.22
#7	85.32%	79.83%	82.48%	0.17
#8	67.91%	96.75%	79.80%	0.26
Mean	77.50%	82.93%	78.82%	0.31
± std	9.61%	14.02%	3.27%	0.16
± CI (95%)	8.42%	12.29%	2.86%	0.14
± CV	12.40%	16.90%	4.14%	0.50

6.3.4.8 Clinical testing and Results on Patients

From the initial fifty-eight tumors of our dataset, eight tumors have been used in the classification training set while fifteen patients with brain metastases (MET PET), twenty-five patients with HNC (FDG PET), and ten patients with lung cancer

(FDG PET) have been considered to assess the differences between our segmentations and the "ground truth" provided by three expert operators [36].
Table 6.3.20, Table 6.3.21 and Table 6.3.22 show sensitivity, PPV, DSC and HD results for brain metastases, HNC, and lung cancers.

Table 6.3.20 Mean sensitivities, PPVs, DSCs and HDs for 15 brain cancer studies using MET-PET examinations are reported. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

Brain Cancer	**Sensitivity [%]**	**PPV [%]**	**DSC [%]**	**HD [voxels]**
#1	94.20%	91.70%	92.90%	0.53
#2	92.30%	87.40%	89.80%	0.59
#3	81.90%	90.30%	85.90%	2.28
#4	87.68%	92.55%	89.70%	0.44
#5	92.73%	91.29%	91.77%	0.50
#6	83.40%	92.80%	87.90%	1.01
#7	91.10%	88.30%	89.70%	1.14
#8	91.90%	91.00%	91.40%	0.46
#9	89.80%	80.60%	85.00%	1.28
#10	92.00%	90.50%	91.20%	1.10
#11	88.50%	88.80%	88.70%	0.98
#12	88.90%	88.10%	88.50%	2.23
#13	90.80%	93.40%	92.10%	1.63
#14	87.60%	86.70%	87.20%	1.39
#15	90.90%	92.90%	91.90%	0.55
Mean	89.58%	89.76%	89.58%	1.07
± std	±3.40%	±3.31%	±2.37%	±0.61
± CI (95%)	±1.72%	±1.67%	±1.20%	±0.31
± CV	±3.80%	±3.69%	±2.64%	±0.57

No distinction between lesion volumes is reported as for phantom sphere results because all considered tumors are greater than 2.5 ml (lesions with a sphere-equivalent diameter greater than 17 mm). In addition, most segmented tumors (~90%) are homogenous (one single peek is visually detected on the histogram of lesion voxel values).
Figure 6.3.9 reports the quantitative comparison between the proposed segmentations and the gold-standards. Automatically segmented BTVs showed high agreement with the manually segmented BTVs (determination coefficient R2 = 0.98).

Table 6.3.21 Mean sensitivities, PPVs, DSCs and HDs for 25 head & neck cancer studies using FDG-PET examinations are reported. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

Neck & head Cancer	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]
#1	93.30%	88.70%	91.00%	1.25
#2	83.30%	87.50%	85.40%	1.00
#3	89.80%	81.60%	85.50%	1.07
#4	85.50%	91.40%	88.30%	0.93
#5	92.40%	83.80%	87.90%	1.00
#6	94.40%	83.40%	88.50%	1.33
#7	87.70%	95.00%	91.20%	1.18
#8	96.80%	82.20%	88.90%	1.72
#9	95.30%	82.80%	88.60%	1.77
#10	91.10%	85.70%	88.30%	1.31
#11	73.10%	90.50%	80.90%	1.00
#12	95.10%	83.10%	88.70%	1.24
#13	91.40%	74.00%	81.80%	1.28
#14	91.40%	84.20%	87.70%	1.00
#15	93.50%	75.90%	83.80%	1.08
#16	77.60%	84.90%	81.10%	1.07
#17	89.80%	86.30%	88.00%	1.14
#18	91.90%	87.00%	89.40%	0.90
#19	82.70%	89.50%	86.00%	1.18
#20	92.60%	89.30%	90.90%	0.57
#21	92.20%	88.60%	90.40%	2.64
#22	90.30%	83.80%	87.00%	0.94
#23	83.00%	90.70%	86.70%	1.00
#24	86.70%	76.50%	81.30%	1.00
#25	91.20%	91.80%	91.50%	0.92
Mean	89.28%	85.53%	87.15%	1.18
± std	±5.70%	±5.13%	±3.23%	±0.39
± CI (95%)	±2.24%	±2.01%	±1.27%	±0.15
± CV	±6.39%	±5.99%	±3.71%	±0.33

Table 6.3.22 Mean Sensitivities, PPVs, DSCs and HDs for 10 lung cancer studies using FDG-PET examinations are reported. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

Lung Cancer	**Sensitivity [%]**	**PPV [%]**	**DSC [%]**	**HD [voxels]**
#1	94.30%	87.40%	90.70%	2.59
#2	90.60%	92.10%	91.40%	1.34
#3	85.70%	91.40%	88.50%	1.82
#4	85.90%	83.80%	84.80%	1.08
#5	89.90%	96.30%	93.00%	1.07
#6	81.70%	80.30%	81.00%	1.66
#7	77.00%	86.50%	81.50%	1.07
#8	91.20%	92.10%	91.60%	0.94
#9	90.70%	85.60%	88.10%	2.14
#10	93.00%	86.40%	89.50%	1.59
Mean	88.00%	88.19%	88.01%	1.53
± std	±5.41%	±4.73%	±4.23%	±0.54
± CI (95%)	±3.35%	±2.93%	±2.62%	±0.33
± CV	±6.14%	±5.37%	±4.80%	±0.35

Table 6.3.23 shows the comparison between the proposed segmentation with respect to the original LAC and RW methods. For sake of simplicity, T42%, RG and FCM methods were not considered in patient studies because they were found to perform consistently worse, in phantom experiments, than the LAC and RW methods.
Analysis of variance (ANOVA) on the DSC were used to test for statistical differences between methods considering all patient lesions (n=50). The proposed method showed a mean DSC of 88.05±3.33%, the RW of 83.73±5.57%, and LAC of 80.40±6.86%. Table 6.3.24 compares the results of ANOVA for the segmentation methods indicating significant differences ($p<0.05$) between methods and, consequently, the proposed segmentation yielded better performance in minimizing the difference between manual and automated BTVs than the other two reference methods. Finally, in order to confirm that the proposed method is operator-independent, a sub-dataset of 5 brain metastases, 5 HNC, and 5 lung cancers was segmented by five different operators (the three experts mentioned above and two non-specialists). The result consisted of five identical BTVs.

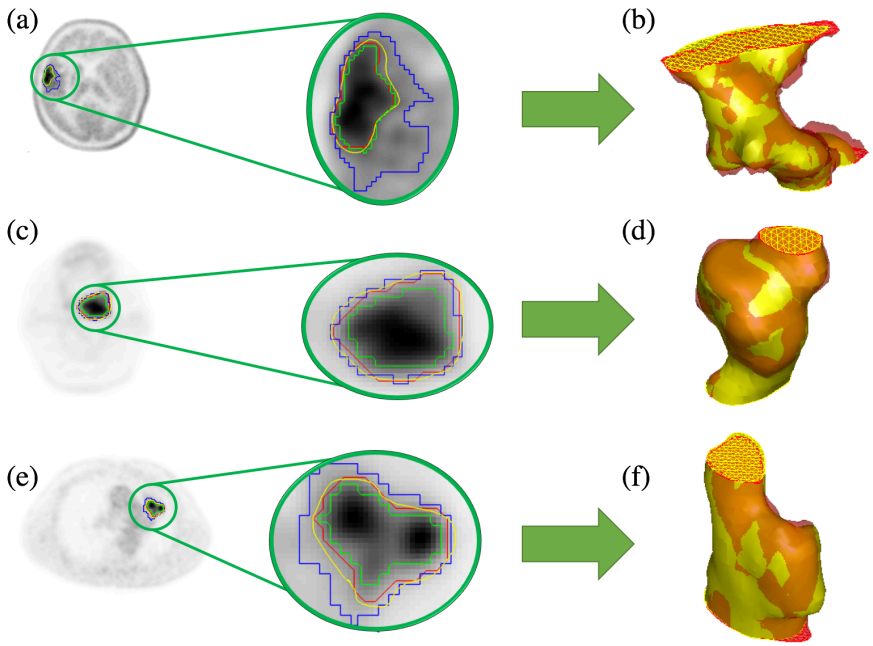


Figure 6.3.9 Three results achieved by the proposed segmentation method on PET images are reported in figure (a, c, e), respectively. The proposed segmentations (red contours) and the gold standards (yellow contours) are superimposed. Blue and green contours concern the DA tissue classification. The region outside the blue boundaries is labelled as background, the region inside the green boundaries is labelled as lesion, and the region between blue and green contours is labelled as undefined. The tridimensional reconstructions of the tumors are shown in figure (b, d, e): manual (yellow) and proposed (red) BTVs are rendered with transparent surfaces to emphasize volume intersections.

Table 6.3.23 Sensitivities, PPVs, DSCs and HDs for BTV segmentation using the proposed, original LAC and RW methods.

Cancer	**Sensitivity** [Mean ± std]	**PPV** [Mean ± std]	**DSC** [Mean ± std]	**HD** [Mean ± std]
		The proposed method		
Lung	88.00 ± 5.41%	88.19 ± 4.73%	88.01 ± 4.23%	1.53 ± 0.54
Head & Neck	89.28 ± 5.70 %	85.53 ± 5.13%	87.15 ± 3.23 %	1.18 ± 0.39
Brain	89.58 ± 3.40%	89.76 ± 3.31%	89.58 ± 2.37%	1.07 ± 0.61
		Original LAC		
Lung	92.68 ± 5.67%	71.92 ± 12.36%	80.33 ± 9.51%	3.29 ± 0.74
Head & Neck	80.32 ± 10.82 %	83.88 ± 12.06 %	78.73 ± 6.68 %	2.81 ± 0.49
Brain	91.16 ± 6.16%	77.38 ± 13.89%	83.55 ± 9.03%	2.88 ± 0.54
		RW		
Lung	92.55 ± 5.66%	74.72 ± 7.44%	82.03 ± 6.52%	2.29 ± 0.77
Head & Neck	79.03 ± 7.29%	84.58 ± 9.57 %	82.59 ± 4.95%	1.41 ± 0.40
Brain	88.12 ± 8.36%	86.68 ± 3.43%	86.75 ± 5.14%	1.31 ± 0.64

Table 6.3.24 ANOVA on the DSC showed statistical differences between segmentation methods.

ANOVA	F value	F critic value	P-value
Proposed Method vs RW	22.1389602	3.93811108	8.3285E-06
Proposed Method vs LAC	50.2234164	3.93811108	2.1246E-10
RW vs LAC	7.06494187	3.93811108	0.00918045

6.3.4.9 Discussion

In this study, an enhanced LAC algorithm has been adapted to the PET imaging field and combined with a DA classifier to obtain operator independent segmentations for RTP purposes and for therapy response assessment.
In order to combine the LAC, which is a gradient descent based minimization method with information derived from learning (i.e. the DA classifier), an innovative energy function is proposed.
The classifier has been purposely trained to label PET tissues into three categories: normal, abnormal, and border-line tissues. The training procedure is based on the PET "ground truth" obtained from manual segmentations provided by an expert clinical staff and processed using proper tools [36]. After this preparation step, the DA classifier is able to label tissues never encountered before. Such useful information is integrated into the LAC algorithm in order to enhance the segmentation performance.
By algorithm construction the BTV delineation is independent of variation in the initial user input. This input is required to avoid any false positive or critical conditions because the FDG distribution is, in general, not limited to cancer area. FDG enters the cells according to glucose transport mechanism. For this reason, active healthy tissues can show high radio-tracer uptake and, consequently, an initial yet limited supervision by the operator is preferred. However, in the case of brain metastases, it is possible to take advantage of the great sensitivity and specificity of MET radio-tracers which is able to discriminate between healthy versus abnormal tissues. Consequently, in MET-PET studies, user input is not necessary and the process can be fully automatic [84]. As a final remark, since an operator-independent mask is created during the initialization step and because a fully automatic stop condition is included, the proposed method produces extremely repeatable results (as confirmed by the inter operator segmentation test in the previous sections). An observation that is worth of note is that the proposed model is not asymmetric in order to preserve the desired contrast direction of the user initialization step. The stopping criterion terminates the algorithm as soon as this established contrast

direction is violated. In short, the algorithm is deliberately designed to capture brighter, rather than darker, regions as explained in the previous sections.

Performance measurements of the proposed method have been obtained by phantom experiments. We are aware that they cannot replicate all the aspects of a real clinical case. Nevertheless, they represent a useful tool to assess common reference performances across different segmentation algorithms. The aim of this test is to evaluate the efficiency of the method and fairly compare its performance with other methods under identical conditions where results may be quantitatively and reliably evaluated. For this purpose, using a well-defined sharp-edged target is preferable because the only uncertainty on the exact boundary location is introduced by the point spread function of the measuring system.

In the case of phantom experiments using plastic spheres, PET image experiments should be carried out in phantoms with inactive background [88] to reduce the cold wall effect caused by the plastic wall that separates the spheres from the background. This model may not be useful to test delineation algorithms because this condition is different from patient studies. For this reason, we insert hot spheres in a warm background. Nevertheless, these spheres are homogeneous targets. In our patient dataset a single peak representing homogeneity is generally detected on the histogram of segmented BTVs. This further motivated the use of phantom with spheres. However, in order to perform an even more robust validation, we included in our workflow the PET acquisition by Soffientini et al [89]. In the latter, they employed zeolites to generate tumor-like inserts that do not suffer from the cold wall effect and possessing any desired shape, size, and contrast levels. Further, zeolites generate PET images with heterogeneous appearance (multiple peaks are detected on their histogram).

We believe that the combined use of these two different class of phantoms can be considered a suitable tool to judge the reliability of our segmentation algorithm. The LAC-DA method showed DSC and sensitivity greater than 90% for the spheres with a diameter greater than 17 mm providing better results than the other state-of-the-art methods. The segmentation of small lesions remains challenging [109]. The PVE cannot be assumed to be negligible when lesion sizes are smaller than 2–3 times the FWHM of the point spread function of the PET image resolution reconstructed by the PET imaging system [77]. Zeolite phantom experiment confirmed optimal results, even in case of irregular and heterogeneous targets, if compared to those presented in Table 3 of the recent study of Berthon et al [111].

In clinical cases, the only valid ground truth for quantitative segmentation evaluation is impossible to obtain because the radio-treatment alters the tumor morphology. Consequently, manual contouring of three experts has been used as a gold-standard although it is a challenging task in deciding which regions to include or exclude in the BTV (i.e., radiotherapy planning experts tend to draw larger contours than nuclear medicine physicians). The tool proposed in [36] allows overcoming such limitations producing a consolidated reference to be used to assess the applicability

of the proposed algorithms in the clinical environment under different conditions (radio-tracers and body districts). Our algorithm has been engineered to be optimal in the clinical practice where the radiotherapist contours a target volume containing the segmented BTV plus a surrounding volume with some significant probability of containing microscopic disease invisible to the PET investigation. Segmentation is optimal for isolated targets, whether requiring single or multiple independent segmentations. Studies on heterogeneous zeolites showed that the algorithm can accommodate the segmentation of multiple small target coexisting in the same VOI. For what concerns the search for post-treatment residual tumor, the same considerations hold. However, in the latter case, the major role is played by the resolution of the PET as compared with the dimension of tumor residues, rather than by the capabilities of the segmentation algorithm itself.
Fifty patients underwent radiotherapy treatment have been considered. Depending on the body district involved, patients underwent FDG or MET PET examinations. Results showed that the proposed method can be considered clinically feasible, since it has been integrated in the current clinical practice. In addition, automatically delineated BTVs showed high agreement with the ground truth (R2 = 0.98).
Finally, the proposed method is different from joint PET/CT or PET/MRI segmentation approaches. One-to-one correspondence between functional and anatomical regions is often unrealistic [86] because different imaging techniques convey different information. For example, the BTV can radically change the RTP, as reported in [112], because radio-tracer uptake was identified in a lymph node which was invisible in the CT images. For this reason, multimodal imaging must be properly evaluated according to the specific clinical scenario [113]. However, CT or MRI studies can be used to identify healthy tissues (i.e. heart, brain, bladder, etc…) and, consequently, to remove false positives from PET analysis.

6.3.5 A Fully Automated Segmentation System of Positron Emission Tomography Studies

6.3.5.1 Phantom Studies

Four phantom experiments were used for preliminary assessment of the performance. The phantom is composed of an elliptical cylinder (minor axis = 24 cm, major axis = 30 cm, h = 21 cm) containing six different spheres (diameters: 10, 13, 17, 22, 28, and 37 mm) placed at 5.5 cm from the center of the phantom. The ratio between sphere and background radioactivity concentration ranged from 3:1 to 8:1. Performances were evaluated by grouping the results with respect to sphere diameters: small spheres, i.e. diameter smaller than 22 mm, and large spheres, with a diameter greater than 17 mm. This choice was motivated by the fact that large biases are introduced by the partial volume effect [77] in PET imaging.

6.3.5.2 Clinical studies

Ten patients with brain metastases were retrospectively considered. These patients were referred to diagnostic PET/CT scan before Gamma Knife (Elekta, Stockholm, Sweden) treatment. Tumor segmentation was performed off-line without actually influencing the treatment protocol or the patient management. No sensitive patient data were accessed. As such, after all patients were properly informed and released their written consent, the institutional hospital medical ethics review board approved the present study protocol. Patients fasted 4 h before the PET examination, and successively were intravenously injected with MET. The PET/CT oncological protocol started 10 min after the injection.

6.3.5.3 PET/CT Acquisition Protocol

All acquisitions in this study were performed within the same hospital department and using the same equipment, a Discovery 690 PET/CT scanner (General Electric Medical Systems, Milwaukee, WI, USA). The PET protocol included a SCOUT scan at 40 mA, a CT scan at 140 keV and 150 mA (10 sec), and 3D PET scans. The 3D ordered subset expectation maximization algorithm was used for the PET imaging. Each PET image consists of 256×256 voxels with a grid spacing of 1.17 mm3 and thickness of 3.27 mm3. Consequently, the size of each voxel is 1.17 × 1.17 × 3.27 mm3. Thanks to the injected PET radio-tracer, tumor appears as hyper-intense region. The CT scan was performed contextually to the PET imaging and used for attenuation correction. Each CT image consists of 512 × 512 voxels with size 1.36 × 1.36 × 3.75 mm3.

6.3.5.4 Framework for performance evaluation

Overlap-based and spatial distance-based metrics are considered to determine the accuracy achieved by the automatic segmentation system against the gold-standard [[101]. In particular, the formulations of dice similarity coefficient (DSC), and Hausdorff distance (HD) are used.

DSC measures the spatial overlap between the reference volume and the segmentation system: a DSC value equal to 100% indicates a perfect match between two volumetric segmentations, while DSC = 0% indicates no overlap. Nevertheless, overlap-based metrics are not well suited for small anomalies. For this reason, distance-based metrics are preferable, especially when the boundary segmentation is critical, such as in BTV delineation for RTP. In particular, HD is used to measure the most mismatched boundary voxels between automatic and manual BTV: small HD

means an accurate segmentation, while a large HD is synonymous of poor accuracy. Finally, the performance of the proposed method is compared to other state of the art BTV segmentation methods: the original LAC method [84], the RG method [96], the enhanced RW method such as described in [86], and the FCM clustering method [109].

6.3.5.5 Phantom Studies

Performance results from phantom experiments were divided considering small and large spheres, in four independent cases, each carried out with different signal ratios. The accuracy improved for all spheres, regardless of their volume, when the signal ratio increased. In general, due to the partial volume effect, the separation of small targets from the background is very challenging, and the difficulty increases in condition of low signal contrast. The sphere volumes are underestimated with more false negatives than false positives. The dice similarity coefficient (DSC) rate is 77.51±3.46% and the Hausdorff distance (HD) is 1.12±0.15 voxels. For the spheres with a diameter greater than 17 mm, excellent performances are obtained with a DSC rate greater than 92% (HD = 1.06±0.09). The mean difference between segmented and actual volumes is positive (the sphere volumes are overestimated); larger margins can help to prevent the extension of tumor infiltration.
The performance of the system was compared to other state of the art PET image segmentation methods. In particular, the original LAC [83], the RG [96], the RW [86], and the FCM [109] methods have been used for comparison. Table 6.3.25 summarizes the results and shows that this automatic segmentation outperforms the methods tested for comparison for all the considered cases.

Table 6.3.25 DSC and HD values for the proposed method and other state of the art PET image segmentation methods.

	DSC	HD (voxels)
Our System	84.79 ± 8.00%	1.09 ± 0.12
Original LAC	82.55 ± 7.56%	1.44 ± 0.55
RW	82.12 ± 8.78%	1.22 ± 0.43
RG	79.01 ± 9.34%	1.67 ± 0.57
FCM	77.13 ± 8.79%	1.68 ± 0.49

6.3.5.6 Clinical Studies

The performance of the presented system is investigated considering ten

metastases against the ground truth provided by three expert operators. In clinical cases, the histopathology analysis is unavailable after the gamma knife treatment. For this reason, the manual delineation performed by expert clinicians is a commonly accepted substitute for ground truth to assess the clinical effectiveness and feasibility of PET delineation methods. Consequently, manual segmentations performed by three experts are used to define a consolidated reference using STAPLE algorithm [36]. This simultaneous ground truth estimation tool combines a collection of segmentations into a single and consolidated ground truth segmentation. It computes a probabilistic estimate of the true segmentation estimating an optimal combination of the segmentations. This algorithm is formulated as an instance of the expectation maximization (EM).

Differently from the phantom studies, no discussion of the tumor volumes is provided here, mostly because all considered BTVs are greater than 2.5 ml (lesions with a sphere-equivalent diameter greater than 17 mm). In particular, tumor volumes ranged from 2.69 ml to 20.49 ml (mean ± std = 7.08 ± 5.81 ml). The ratio between lesion and background radioactivity concentration ranged from 2.76:1 to 7.40:1 (mean ± std = 3.88:1 ± 1.45:1). These values are included in the range of the phantom experiments used for preliminary performance testing. For this reason, although phantom studies don't replicate all the properties of real lesions, they represent a useful tool to assess performances across different segmentation methods.

Table 6.3.26 summarizes the comparison between this automatic segmentation and the original LAC and RW approaches. Since LAC and RW outperformed RG, and FCM methods on the phantom experiments, the latter algorithms were not considered in patient studies. The automatic algorithm performed better than LAC and RW methods minimizing the difference between references and automated BTVs.

Figure 6.3.10 reports the comparison between the proposed segmentations and the gold-standards.

Table 6.3.26 DSCs and HDs using the proposed system, original LAC and RW methods.

	DSC	HD
Our System	88.35 ± 2.60%	1.42 ± 0.57
Original LAC	83.77 ± 8.53%	2.97 ± 0.68
RW	87.01 ± 5.16%	1.38 ± 0.74

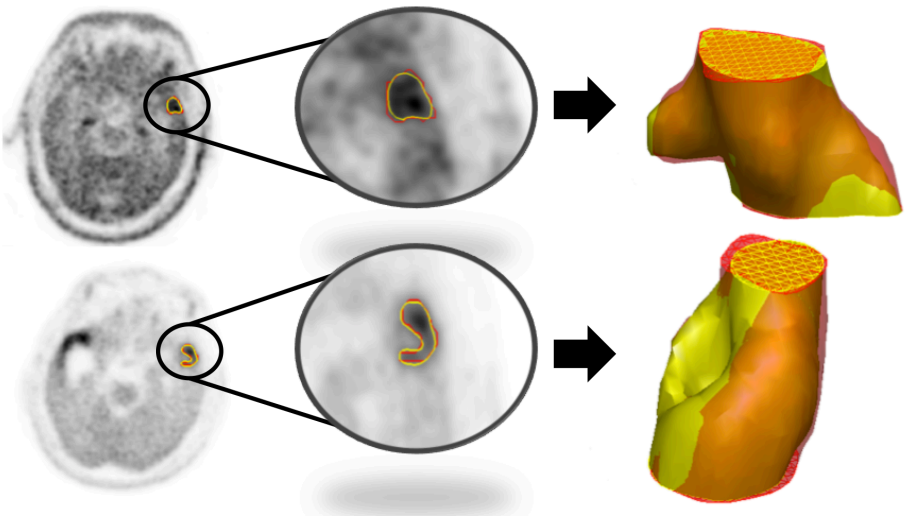


Figure 6.3.10 Examples of automatic segmentations. The retrieved segmentations and the gold standards are shown in red and yellow, respectively.

6.3.6 Tissue Classification to Support Local Active Delineation of Brain Tumors

6.3.6.1 Dataset

Seventeen patients with brain metastases referred to diagnostic MET PET/CT scan have been retrospectively considered. The scan only interested the brain region. Patients fasted 4 h before the PET examination, and successively were intravenously injected with MET. The PET/CT oncological protocol started 10 min after the injection. Tumor segmentation was performed off-line without influencing the treatment protocol or the patient management. The institutional hospital medical ethics review board approved the study protocol and all patients involved were properly informed released their written consent.

6.3.6.2 PET/CT Acquisition Protocol

PET/CT scans were performed using the Discovery 690 PET/CT scanner (General Electric Medical Systems, Milwaukee, WI, USA). The PET protocol includes a SCOUT scan at 40 mA, a CT scan at 140 keV and 150 mA (10 sec), and 3D PET scans. The 3D ordered subset expectation maximization algorithm was used to for the PET imaging. Each PET image obtained consists of 256×256 voxels with a grid spacing of 1.17 mm3 and thickness of 3.27 mm3. Consequently, the size of each voxel is 1.17 × 1.17 × 3.27 mm3. Thanks to the injected PET radiotracer, tumor

appears as hyper-intense region. The CT scan was performed contextually to the PET imaging and used for attenuation correction. Each CT image consists of 512 × 512 voxels with size 1.36 × 1.36 × 3.75 mm3.

6.3.6.3 Framework for performance evaluation

A framework for the evaluation of the proposed protocol is presented. The effectiveness of the tissue classification is calculated regarding correct/incorrect classification using sensitivity, specificity, precision, and accuracy scores. Overlap-based and spatial distance-based metrics are considered to determine the accuracy achieved by the proposed computer-assisted segmentation system against the gold-standard (i.e., the manual segmentations performed by three experts are used to define a consolidated reference as described in the next section).
The sensitivity is the number of correctly classified positive samples divided by the number of true positive samples, while the specificity is the number of correctly classified negative samples divided by the number of true negative samples. The precision is related to reproducibility and repeatability and it is defined as the degree to which repeated classifications under unchanged conditions show the same results. The accuracy is defined as the number of correctly classified samples divided by the number of classified samples. Concerning segmentation algorithm performance, the formulations proposed in [110] are used. In particular, mean, and standard deviation of sensitivity, positive predictive value (PPV), dice similarity coefficient (DSC), and Hausdorff distance (HD) were calculated.

6.3.6.4 Gold Standard

The ground truth requires exact knowledge of the tumor and the histopathology analysis provides the only valid ground truth for the PET quantitative assessment. Nevertheless, the histopathology analysis is unavailable after the treatment. For this reason, the manual delineations performed by expert clinicians are a commonly accepted substitute for ground truth to assess the clinical effectiveness and feasibility of PET delineation methods [78]. Nevertheless, manual delineation is often influenced by the clinical specialization of the operator. For example, oncologists will, on average, draw smaller BTVs than radio-therapists. For this reason, the segmentations performed by three experts with different expertise (the chief nuclear medicine physician –M.I. author-, the chief radiotherapy physician –M.S. author- and an expert radiotherapy physician –G.R. author) were used as “ground truths”. A simultaneous ground truth estimation tool [36] was employed, and the three segmentations were combined to define a single and consolidated ground truth for each study.

6.3.6.5 Classifier validation

The optimal K value of the K-Fold cross-validation integrated in the classifier has been determined as 5 through the trial-and-error method (k range: 5–15, step size of 5) [56][65]. It corresponds to the highest classification accuracy. The validation results are shown in Table 6.3.27.

Table 6.3.27 Sensitivity, specificity, precision, and accuracy values for KNN, NB, and DA classifier validations.

	Sensitivity	Specificity	Precision	Accuracy
KNN	97.13%	81.54%	98.88%	95.25%
NB	87.09%	94.94%	99.68%	86.73%
DA	90.00%	85.12%	99.02%	88.31%

6.3.6.6 Clinical testing and Results on Dataset

Seventeen brain lesions were considered. From the initial dataset, two tumors were used in the classifier training. Consequently, the performance of the presented algorithm is investigated in the remaining cases against the ground truth provided by three expert operators are depicted in Table 6.3.28.

Table 6.3.28 Mean sensitivities, PPVs, DSCs and HDs for 15 lesions are reported to assess the differences between the segmentations obtained using the LAC method with KNN, NB and DA classifiers and the "ground truth" provided by the three operators.

	Sensitivity (%)	PPV (%)	DSC (%)	HD (voxels)
KNN	91.54±1.35	85.36±3.72	88.27±1.91	1.18±0.52
NB	89.58±3.40	89.76±3.31	89.58±2.37	1.07±0.61
DA	91.21±1.93	90.78±2.03	90.92±1.35	0.79±0.40

In addition, regardless of the classifier used, automatically segmented BTVs show high agreement with the manually segmented BTVs (determination coefficient R2 = 0.98). Figure 6.3.11 reports the comparison between the proposed segmentations and the gold-standards for two patients.

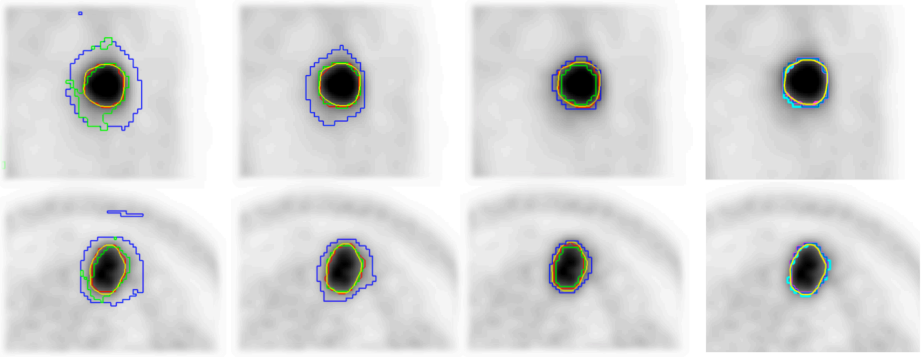


Figure 6.3.11 Two example of brain tumor segmentation for each row using the LAC method coupled with KNN (first column), NB (second column), and DA (third column) classifiers. The proposed segmentations (red contours) and the gold standards (yellow contours) are superimposed. Blue and green contours concern the tissue classification. The region outside the blue boundary represent the "background", while the region inside the green boundaries, and between the curves, represent the "lesion", and the "border-line" region, respectively. In the last column, all retrieved segmentations are superimposed (gold standard in yellow, KNN in magenta, NB in cyan, and DA in light blue).

6.3.6.7 Discussion

In this study, we described a segmentation protocol which leverages on the properties of MET PET to achieve the fully automatic segmentation of brain cancer. In this context, we used a segmentation algorithm LAC which features an energy function specifically adapted to the PET imaging and which combines the information from the PET data, and feedback from a classifier. The aim of such a protocol is for radiotherapy treatment purpose and for therapy response assessment. Each classifier was purposely and independently trained to label PET tissues into normal, abnormal, and border-line tissues. The training procedure is based on the ground truth obtained using a proper tool [36] starting from the manual segmentations provided by three expert operators. In addition, a smart sampling operation based on a moving window of 3x3 voxels has been implemented. After this preparation step, each classifier was able to label tissues never encountered before and to convey this useful information into the LAC algorithm. As a final remark, a fully automatic stop condition was provided. In this way and by taking advantage of the great sensitivity and specificity of MET PET studies to identify tumors in brain area, the proposed system produces segmentation results which are completely independent by the user.

Performance of the proposed method were obtained by patient studies, for which the ground is impossible to obtain because the treatment alters the tissue morphology. Consequently, although manual delineation may largely differ between different human operators (for example, radiotherapy experts tend to draw larger boundaries than nuclear medicine physicians), with obvious impact on the resulting surrogate of ground truth, the delineation from experts is the only alternative. Here, for

performance evaluation purposes, the manual segmentations from three experts was used as a gold-standard. Such manual delineations were used to produce a consolidated reference [36] which was then used to assess the feasibility of the proposed method in the clinical environment. Seventeen patients undergo a PET/CT scan have been considered. Results show that the proposed protocol can be considered clinically feasible using any of the tested classifier, although the DA delivered slightly better results. Automatically segmented BTVs showed high agreement with the gold-standard (R2 = 0.98). Considering the relevance of our study in oncological patient management we will add PET studies to expand our database and, consequently, to improve and further assess the proposed method.

6.3.7 Comparison of results

Table 6.3.29 shows a comparison of the volumetric accuracy results for the smaller and larger spheres using the proposed segmentation systems 6.3.1, 6.3.3 and 6.3.4. for smaller and larger spheres. The best segmentation systems are respectively 6.3.4 with Sensitivity 75.45±4.02%, PPV 86.89±12.04%, DSC 80.29±6.07%, HD 1.03±0.16 and 6.3.1 with Sensitivity 93.22±7.18%, PPV 87.36±5.09%, DSC 90.67±4.07%, HD 1.18±0.28.

Table 6.3.29 Sensitivity, PPV, DSC, HD rates obtained over NEMA IEC phantoms (10-17mm and 22-73mm) using the proposed system 6.3.1, 6.3.3 and 6.3.4 has been compared.

Sphere Diameter	System	NEMA IEC Phantom	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]
10-17mm	**6.3.1**	Mean± std	70.00±3.49%	87.56±8.87%	77.51±3.46%	1.12±0.15
	6.3.3	Mean± std	74,41±5,98%	85,83±13,94%	78,84±6,18%	1,06±0,19
	6.3.4	Mean± std	75.45±4.02%	86.89±12.04%	80.29±6.07%	1.03±0.16
22-37mm	**6.3.1**	Mean± std	93.22±7.18%	87.36±5.09%	90.67±4.07%	1.18±0.28
	6.3.3	Mean± std	92,58±2,49%	88,45±4,48%	90,36±2,75%	1,20±0,20
	6.3.4	Mean± std	92.27±2.45%	87.98±4.84%	90.04±3.23%	1.26±0.30

Table 6.3.30 shows the comparison between the proposed segmentation systems 6.3.1, 6.3.3 and 6.3.4 with respect to the original LAC and RW methods. In clinical cases the best segmentation system is 6.3.4 with DSC 85.36 ± 2.94%, 85.98 ± 3.40%, 88.02 ± 2.75% respectively for lung, Head & Neck and Brain.

Table 6.3.30 Sensitivities, PPVs, DSCs and HDs for cancer studies using the proposed system 6.3.1, 6.3.3, 6.3.4, original LAC and RW methods has been compared.

Cancer	**Sensitivity** [Mean ± std]	**PPV** [Mean ± std]	**DSC** [Mean ± std]	**HD** [Mean ± std]
	6.3.1 System			
Lung	93.63 ± 4.55%	78.64 ± 4.41%	85.36 ± 2.94%	1.87 ± 0.62
Head & Neck	91.00 ± 7.33%	82.37 ± 7.30%	85.98 ± 3.40%	1.39 ± 0.61
Brain	90.17 ± 4.89%	86.60 ± 6.89%	88.02 ± 2.75%	1.28 ± 0.54
	6.3.3 System			
Lung	88.09 ± 4.18%	87.84 ± 5.15%	87.88 ± 3.89%	1.73 ± 0.74
Head & Neck	88.58 ± 6.33 %	86.30 ± 5.37%	87.14 ± 3.31 %	1.16 ± 0.29
Brain	92.33 ± 1.64%	86.43 ± 3.72%	89.21 ± 2.17%	1.08 ± 0.52
	6.3.4 System			
Lung	88.00 ± 5.41%	88.19 ± 4.73%	88.01 ± 4.23%	1.53 ± 0.54
Head & Neck	89.28 ± 5.70 %	85.53 ± 5.13%	87.15 ± 3.23 %	1.18 ± 0.39
Brain	89.58 ± 3.40%	89.76 ± 3.31%	89.58 ± 2.37%	1.07 ± 0.61
	Original LAC			
Lung	92.68 ± 5.67%	71.92 ± 12.36%	80.33 ± 9.51%	3.29 ± 0.74
Head & Neck	80.32 ± 10.82%	83.88 ± 12.06%	78.73 ± 6.68%	2.81 ± 0.49
Brain	91.16 ± 6.16%	77.38 ± 13.89%	83.55 ± 9.03%	2.88 ± 0.54
	RW			
Lung	92.55 ± 5.66%	74.72 ± 7.44%	82.03 ± 6.52%	2.29 ± 0.77
Head & Neck	79.03 ± 7.29%	84.58 ± 9.57%	82.59 ± 4.95%	1.41 ± 0.40
Brain	88.12 ± 8.36%	86.68 ± 3.43%	86.75 ± 5.14%	1.31 ± 0.64

CHAPTER 7

7 Discussions and Conclusions

PET segmentation in radiotherapy is a critical task due to the lack of consistency in tumor contour, low image resolution, and relatively high level of noise and heterogeneity of FDG uptake within a lesion. Nevertheless, accurate lesion segmentation in PET imaging is essential for an accurate quantification of prognosis assessment and therapy response. Qualitative visual interpretation is the most commonly used method. The manual segmentation method is dependent on the experience of the nuclear physician limiting the measurement accuracy. In addition, it is time-consuming and impaired by inter- and intra-observer variability that is due to subjectivity and sensitivity to the display window level settings. To reduce these issues, several automatic methods have been presented but few clinical studies are available and there is no consensus for proper BTV determination. To date, CT and MR imaging are considered to be the standard for target volume delineation in radiotherapy. On the other hand, CT and MR imaging does not show the biological features of tumors. For this reason, PET has been introduced in the radiotherapy field to assist radiation oncologists in clinical routine. Despite the fact that many studies use the information of co-registered PET and CT images to identify features for distinguishing a lesion from the background and, consequently, for PET image segmentation, I believe that BTV extraction is independent from anatomical segmentation being inappropriate to consider a matching between anatomical and metabolic regions, with special reference in HNC district. I believe that a correct delineation of BTV must be obtained without incorporating anatomical information or incorporating them with great attention. For example, MRI is crucial for the illustration of the anatomy of brain tumors: for this reason, a study to determine simultaneously tumor contours on PET and RM images is a topic of this thesis. In addition, to identify oncological lesions in whole-body PET scans, a completely automatic detection method cannot be implemented since healthy organs such as brain, heart, bladder, and kidneys normally have a high FDG uptake, and, consequently, user interaction is mandatory. Vice versa, this is feasible in brain MET-PET studies.

In this ph.D. Thesis a series of contributions concerning medical image processing and segmentation, and data analysis to classify and delineate tumor 3D volume to support medical decision have been presented.

An automatic multi-seed detection method for magnetic resonance (MR) breast image segmentation is presented. The multi-seed detection has been focused because of its importance in regional segmentation technique as the region growing. The

maximum concavity points have been proposed as the seed points for the breast MR image segmentation. The detection of these points is based on the identification of three ROI of the breast MR image: the axillary regions and the sternal region. The Gold Standard, described in experimental results section, is used to compute the effectiveness and the performance of the proposed method. The preliminary results are very encouraging in terms of statistical metrics and execution time. In future works we are interested to extend the number of cases study and to develop a CAD (computer aided diagnostic) to detect suspicious regions on breast MRI: the first step is to detect the region of interest by using our proposed method (segmentation phase), than a further analysis and investigation should be conducted to detect suspicious regions by analysing several features such as texture descriptors, statistical descriptors, histogram of gradients and others state of the art techniques.

A KSVM based technique for CD affected human patients classification has been presented. The classification results on the UPPH dataset composed of 300 patients, each of one codified by 22 qualitative features, have been calculated and compared against the histological specimen results. Histological specimen results are the clinical Ground-Truth for CD diagnosis and have been used for demonstrating the validity of proposed technique.

Sensitivity, Specificity, Negative Predictive Value, Precision, Accuracy, and Error scores obtained using the proposed technique highlight the improvement in accuracy and quality compared to the manual results presented in [7].

The proposed technique has a significant advantage in integrating the cross-validation strategy. In addiction, the cross-validation strategy gives its contribution to the classifier generalization capability without showing either over-fitting or sensibility to the selected test dataset. However, classifier usability does not require any parameter setting and deep knowledge about the used learning machine technique. So, its degree of acceptance in medical practices is very high.

It was developed a KSVM-based system to classify incisional hernia recurrence. The classification results on a UPPH Dataset of 154 vectors composed of 34 qualitative characteristics extracted from expert surgeons were compared with the results of the data extrapolated from the clinical records. Sensitivity (86,25%), Specificity (87,14%), Negative Predictive Value (84,72%), Precision (88,46%), Accuracy (86,67%), and Error (13,33%) scores obtained using the proposed technique highlight the validity for the relapse's classification methodology Table 6.2.2.

The intelligent data analysis therefore confirmed the already obvious correlations showed in the international literature, it also revealed possible associations with other comorbidities such as thyroid, liver, genitourinary and other interventions with the development of incisional hernia recurrence. The type of prosthesis and the site of its implant also play a significant role in the development of the recurrence.

For the little number of patients considered, the low data collection and the high number of factors considered in the study despite the discreet sensitivity and

specificity achieved, new studies are required to further investigate new correlations proposals.

Several methodologies have been combined to obtain a system for the smart segmentation of PET volumes. The key advantages of the presented system are that:

- segmentation is performed after converting the PET data in SUVs, which convey patient-specific functional information;
- based on minimal user input (i.e. a rough manual delineation on just one PET slice), a pre-segmentation step identifies the most relevant PET slice and provides an initial properly localized contouring to be used for a subsequent segmentation using LAC;
- a more detailed segmentation process based on LAC and performed following a slice-by-slice based approach provides the flexibility necessary to segment cancer with very irregular shapes;
- an automatic stopping condition is implemented.

As a result, the presented algorithm produces cancer segmentations that are completely independent from variations in the input provided by the user. Such an input is minimal and, since subsequent processing steps are fully automatized, no further interaction is required. Inter-operator variations in the initial input still result in identical segmentations.

Phantom studies were used to verify the effectiveness of the algorithm on well-known, sharp-edged targets. Comparison of the results with the output of multiple alternative algorithms was performed. A large dataset of clinical cases was used to investigate the performance of the system as compared with selected, state of the art alternatives. Such comparison was performed on a statistical basis commonly used and considered as a reference practice in the field. The overall and final conclusion is that this system could easily comply with the demands of everyday clinical activity and could be used to extract in-vivo biomarkers of cancer in treatment response evaluation or to enhance the dose delivery in radiotherapy treatment, in order to avoid cancer recurrence.

A supervised system for normal and abnormal region classification in PET oncological images is presented. Classification results on the data- set composed of 80 patients are calculated and compared against the physician classification.

Sensitivity, specificity, precision, accuracy, and overlap scores obtained using the proposed method highlight the improvement in accuracy and quality compared to the others five methods.

The proposed real-time system could be applied to the field of PET image classification to assist physicians in radiation treatment planning in order to dis-criminate between normal and abnormal tissues.

It was proposed a segmentation system specifically engineered to reach the maximum level of automation and capable of obtaining an operator independent BTV. The segmentation algorithm at the core of the system is a local active contour

method adapted to PET imaging and whose design is based on the minimization of an innovative energy function which incorporates the information from a KNN classifier. The classifier, trained by examples provided by expert human operators, is capable of labelling the tissues into three categories: lesion, background and border-line; potentially incorporating some of the trainer's wisdom in the segmentation process. Considering the strategy adopted to generate the samples for training the classifier, we expected well separated classes. Therefore, the KNN represented the simplest and, to our opinion, the most advantageous choice. Of course, a different classifier might be used as well, and comparison with alternative algorithms will be discussed in future studies. The key features that make our approach original can be summarized as follows: 1) leverage on patient-specific metabolic information through the conversion of PET data into SUV; 2) reduction of the user intervention to minimum and limited to the selection of an initial region of interest on just one PET slice; 3) integration of the tissue classification in the energy functional which drives the local active contour method; 4) a slice-by-slice marching approach, which provides the flexibility necessary to segment cancer with very irregular shapes; 5) an automatic termination condition; 6) a smart automatic process that produces BTVs independent from the initial input. Phantom experiments demonstrated the effectiveness of the method on well-known, sharp-edged targets. The proposed segmentation method showed DSC and sensitivity greater than 90% for the experiments regarding spheres with a diameter greater than 17 mm, and provided better results in minimizing the difference between actual and automated BTVs than the other state-of-the-art methods. Nevertheless, the under-estimation of uptake concentration due to partial volume effect cannot be assumed to be negligible when the lesion size is smaller than 2–3 times the FWHM of the point spread function associated to the PET imaging system. Because of this inherent limitation of PET imaging, the segmentation of small lesions still remains very challenging [55] [96] no matter of the segmentation approach being used. Concerning the clinical application, cases from fifty patients and involving different body districts were considered for performance evaluation and results show that the proposed approach can be considered clinically feasible. Further, produced BTVs showed high agreement with the gold-standard ($R2 = 0.98$). In addition, comparison of the results with the output of state-of-the-art algorithms was performed. Such evaluation was performed on a statistical basis commonly considered to be a reference practice in the bio-imaging field. The overall and final conclusion is that our enhanced LAC method could be used in clinical environment to extract PET biomarkers in therapy response evaluation or to enhance the dose delivery in PTV to avoid cancer recurrence.

The LAC-DA method for the segmentation of biological tumor volumes has been proposed. The key features of the method are:

- The integration of patient-specific functional information obtained by converting PET images to SUV images;

- The inclusion of an initialization step to translate the region highlighted by the user (on just one slice) to an operator independent mask to be used as starting point of the segmentation;
- The integration of the DA classification in the LAC method trough a novel formulation of the energy functional to be minimized;
- The implementation of an automatic termination condition;
- The implementation of a smart automatic process able to segment independently of the user input variations.
- The whole process is nearly full automatized.

Phantom studies were used to verify the performance of the method on well-known, sharp-edged targets. Comparison of the results with the output of state-of-the-art methods was performed. Fifty clinical cases were used to assess the clinical feasibility of the proposed method. Such evaluation was performed on a statistical basis commonly considered as a reference practice in the PET image segmentation.
In conclusion, the proposed LAC-DA method could be used in the clinical environment to extract PET parameters for therapy response evaluation purpose or to enhance the dose delivery in PTV avoiding tumor recurrence.

A complex, semi-automatic system featuring an enhanced LAC algorithm purposely adapted to the PET imaging was further adapted to achieve the fully automatic BTV segmentations of brain cancers. The fully automatic approach leverages on the fact that MET-PET is capable of selectively highlight the ill regions of the brain, so avoiding false positives commonly encountered in other anatomic regions (e.g. as in FDG-PET studies). An automatic and operator-independent ROI is generated around the tumor(s) and used as input to an enhanced LAC algorithm. Then, the LAC performs the BTV delineation. The BTV is built by a slice-by-slice marching approach where the segmentation is performed on subsequent slices. In principle, segmentation through the evolution of a full 3D surface would be preferable. Indeed, while on the one hand we are currently investigating such a 3D approach, on the other hand, the present work moves an important step toward 3D data segmentation improving upon the model proposed by Lankton et al [14] [83] considering the issue of the PET slices thickness (3.27 mm3) far greater than planar resolution (1.17 mm3) which partially justifies the 2D approach. As a final remark, a fully automatic stop condition is provided. In this way, the proposed system produces segmentation results that are completely independent by the user.
Performance of the automatic system has been obtained by phantom studies consisting of hot spheres in a warm background. Nevertheless, phantom experiments cannot replicate all the aspects of a real clinical case but they represent a useful way to assess common performances across different algorithms. DSC greater than 92% for the larger spheres confirm better results in minimizing the difference between reference and automated BTVs than the other state-of-the-art algorithms. We would like to emphasize that original algorithms for both enhanced RW and original LAC

methods [17, 19] [83][86] were optimized for MET-PET brain metastases [10, 12] [84][98]. Concerning RG and FCM methods, we used the source codes available on the web, and we adapted it to our PET dataset.

In patient studies, since radiotherapy treatment alters the cancer morphology over time, histopathology cannot provide a reliable ground truth. Consequently, manual delineation by experts, although it may differ between operators (for example, radiotherapy experts tend to draw larger boundaries than nuclear medicine physicians), is often used as surrogate gold-standard. In this study, we used manual delineations from three experts. To overcome the issue of differences in the manual delineations, a consolidated reference was built [16] [36] and then used to assess the feasibility of the automatic segmentation. PET/CT data from ten patients before Gamma Knife treatment were considered. Results show that the proposed approach can be considered clinically feasible and could be used to extract PET parameters for therapy response evaluation purpose and to assist the BTV delineation during stereotactic radiosurgery treatment planning avoiding cancer recurrence. Finally, further investigations will be carried out to assess the usefulness to introduce in the segmentation a PET tissue classifier capable of influencing the local active contour toward what would be the segmentation performed by a human operator [21–24] [65][93][127][143].

At last, we demonstrated how a semi-automatic algorithm we proposed in previous work may be integrated into a protocol which becomes fully automatic for the detection brain metastases. Such a protocol combines 11C-labeled Methionine PET acquisition with our previous segmentation approach. The key features of the protocol include the use of MET radio tracer, the integration of patient-specific functional information obtained by converting PET images to SUV images; a preliminary task to identify an initial, properly localized, operator-independent ROI to be used for the LAC segmentation; the integration of tissue classification (either using KNN, NB, or DA) in the LAC method directly in the formulation of the energy functional to be minimized to enhance the accuracy of the BTV contouring; and a slice-by-slice marching approach with an automatic termination condition. This features make the whole process fully automatic in the context of brain cancer.

Brain metastases were used to assess the performance of the proposed protocol on a statistical basis commonly considered as a reference practice in the PET imaging field. Results showed that this protocol can be used to extract PET parameters for therapy response evaluation purpose and to provide automatic the BTV delineation during radiosurgery treatment planning in the special context of brain cancer.

In conclusion, the proposed methods are very powerful in terms of PET image and PET/MRI image segmentation, and may be used daily as a Medical Decision Support Systems to enhance the current methodology performed by healthcare operators in radiotherapy treatments. Future directions will aim to investigate smart techniques for the medical imaging problematics, but also for the managing and elaboration of Big Data coming from a wide kind of heterogeneous sources, bringing the computer

science field increasingly closer and transversal to the medicine science.

REFERENCES

[1] Spieth, Spieth, M.E. & Kasner, D.L., 2002. A Tabulated Summary of the FDG PET Literature. Journal of nuclear medicine : official publication, Society of Nuclear Medicine, 43(3), pp.439–441.

[2] Guido, A. et al., 2009. Combined 18F-FDG-PET/CT Imaging in Radiotherapy Target Delineation for Head-and-Neck Cancer. International Journal of Radiation Oncology Biology Physics, 73(3), pp.759–763.

[3] Wahl RL, Jacene H, Kasamon Y, Lodge MA (2009) From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med 50 Suppl 1:122S–50S. doi: 10.2967/jnumed.108.057307

[4] Schwartz, D.L. et al., 2005. FDG-PET/CT imaging for preradiotherapy staging of head-and-neck squamous cell carcinoma. International journal of radiation oncology, biology, physics, 61(1), pp.129–36.

[5] Lauve, A. et al., 2004. Simultaneous integrated boost intensity-modulated radiotherapy for locally advanced head-and-neck squamous cell carcinomas: II--clinical results.

[6] Ciernik, I.F. et al., 2003. Radiation treatment planning with an integrated positron emission and computer tomography (PET/CT): a feasibility study. International Journal of Radiation Oncology*Biology*Physics, 57(3), pp.853–863.

[7] Zaidi, H. et al., 2002. Fuzzy clustering-based segmented attenuation correction in whole-body PET imaging. Physics in medicine and biology, 47(7), pp.1143–1160.

[8] Li, H. et al., 2008. A novel PET tumor delineation method based on adaptive regiongrowing and dual-front active contours. Medical Physics, 35, pp.3711–3721.

[9] Geets, X. et al., 2007. A gradient-based method for segmenting FDG-PET images: methodology and validation. European Journal of Nuclear Medicine and Molecular Imaging, 34, pp.1427–1438.

[10] Schinagl, D.A.X. et al., 2007. Comparison of five segmentation tools for 18 FFLUORO-DEOXYGLUCOSE-POSITRON emission tomography-based target volume definition in head and neck cancer. International Journal of Radiation Oncology Biology Physics, 69, pp.1282–1289.

[11] (I-11)(4-13)(6-14)(7-10)(9-7)Zaidi H, El Naqa I (2010) PET-guided delineation of radiation therapy treatment volumes: A survey of image segmentation techniques. Eur J Nucl Med Mol Imaging 37:2165–2187. doi: 10.1007/s00259-010-1423-3

[12] Ford, E.C. et al., 2006. Tumor delineation using PET in head and neck cancers: threshold contouring and lesion volumes. Med Phys, 33(11), pp.4280–4288.

[13] Drever, L. et al., 2006. A local contrast based approach to threshold segmentation for PET target volume delineation. Medical Physics, 33(6), p.1583.

[14] Drever, L. et al., 2007. Iterative threshold segmentation for PET target volume delineation. Med Phys, 34(4), pp.1253–1265.

[15] Kao C-H, Hsieh T-C, Yu C-Y, et al (2010) F-18-FDG PET/CT-based gross tumor volume definition for radiotherapy in head and neck Cancer: a correlation study between suitable uptake value threshold and tumor parameters. Radiat Oncol. doi: 10.1186/1748-717x-5-76

[16] Paulino, A.C. et al., 2005. Comparison of CT- and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer. International Journal of Radiation Oncology Biology Physics, 61, pp.1385–1392.

[17] Hatt, M. et al., 2009. A Fuzzy Locally Adaptive Bayesian Segmentation Approach for Volume Determination in PET. Ieee Transactions on Medical Imaging, 28, pp.881–893.

[18] (I-18)(4-24)(7-22)Wanet M, Lee JA, Weynand B, et al. (2011) Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: A comparison with threshold-based approaches, CT and surgical specimens. Radiother Oncol 98:117–125. doi: 10.1016/j.radonc.2010.10.006

[19] El Naqa, I. et al., 2007. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning. Medical physics, 34(2007), pp.4738–4749.

[20] H. Yu, C. Caldwell, K. Mah, and D. Mozeg, "Coregistered fdg pet/ct-based textural characterization of head and neck cancer for radiation treatment planning," IEEE Trans. Med. Imag. 28, 374–383 (2009).

[21] Han, D. et al., 2011. Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 245–256.

[22] Song, Q. et al., 2013. Optimal Co-segmentation of tumor in PET-CT images with 87 context information. IEEE Transactions on Medical Imaging, 32(9), pp.1685–1697.

[23] Xia, Y. et al., 2012. Dual-modality brain PET-CT image segmentation based on adaptive use of functional and anatomical information. Computerized Medical Imaging and Graphics, 36(1), pp.47–53.

[24] Potesil, V., Huang, X. & Zhou, X.S., 2007. Automated tumor delineation using joint PET/CT information. In Progress in Biomedical Optics and Imaging - Proceedings of SPIE. p. 65142Y–65142Y–8.

[25] Yezzi A, Zollei L, Kapur T A variational framework for joint segmentation and registration. In: Proc. IEEE Work. Math. Methods Biomed. Image Anal.

(MMBIA 2001). IEEE Comput. Soc, pp 44–51

[26] Leibfarth, S. et al., 2015. Automatic delineation of tumor volumes by cosegmentation of combined PET/MR data. Physics in Medicine and Biology, 60(14), pp.5399–5412.

[27] Bagci U, Udupa JK, Mendhiratta N, et al. (2013) Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images. Med Image Anal 17:929–945. doi: 10.1016/j.media.2013.05.004

[28] Brooksby, B., Pogue, B.W., Jiang, S., Dehghani, H., Srinivasan, S., Kogel, C., Tosteson, T.D., Weaver, J., Poplack, S.P., Paulsen, K.D.: Imaging breast adipose and fibroglandular tissue molecular signatures by using hybrid MRI-guided nearinfrared spectral tomography. Proc. Natl. Acad. Sci. 103(23), 8828–8833 (2006)

[29] Patel, B.C., Sinha, G.: An adaptive k-means clustering algorithm for breast image segmentation. Int. J. Comput. Appl. 10(4), 35–38 (2010)

[30] Lin, M., Chen, J.H., Wang, X., Chan, S., Chen, S., Su, M.Y.: Template-basedautomatic breast segmentation on MRI by excluding the chest region. Med. Phys. 40(12), 122301 (2013)

[31] Wang, L., Chitiboi, T., Meine, H., GÅNunther, M., Hahn, H.K.: Principles and methods for automatic and semi-automatic tissue segmentation in MRI data. Magn. Reson. Mater. Phys. Biol. Med. 29(2), 95–110 (2016)

[32] Wang, L., Platel, B., Ivanovskaya, T., Harz, M., Hahn, H.K.: Fully automatic breast segmentation in 3D breast MRI. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1024–1027. IEEE (2012)

[33] Khazen, M.,Warren, R.M., Boggis, C.R., Bryant, E.C., Reed, S.,Warsi, I., Pointon, L.J., Kwan-Lim, G.E., Thompson, D., Eeles, R., et al.: A pilot study of compositional analysis of the breast and estimation of breast mammographic density using three-dimensional T1-weighted magnetic resonance imaging. Cancer Epidemiol. Prev. Biomark. 17(9), 2268–2274 (2008)

[34] Lee, N.A., Rusinek, H., Weinreb, J., Chandra, R., Toth, H., Singer, C., Newstead, G.: Fatty and fibroglandular tissue volumes in the breasts of women 20–83 years old: comparison of X-ray mammography and computer-assisted MR imaging. AJR. Am. J. Roentgenol. 168(2), 501–506 (1997)

[35] Rosset, A., Spadola, L., Ratib, O.: Osirix: an open-source software for navigating in multidimensional DICOM images. J. digit. Imaging 17(3), 205–216 (2004)

[36] Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging 23(7), 903–921 (2004)

[37] Adams, R., Bischof, L.: Seeded region growing. IEEE Trans. Pattern Anal. Mach. Intell. 16(6), 641–647 (1994)

[38] McKay A, Dixon E, Bathe O, Sutherland F. Umbilical hernia repair in the presence of cirrhosis and ascites: results of a survey and review of the literature. Hernia. 2009 Oct;13(5):461-8.

[39] Calaluce R, Davis JW, Bachman SL, Gubin MM, Brown JA, Magee JD, Loy TS, Ramshaw BJ, Atasoy U. Incisional hernia recurrence through genomic profiling: a pilot study. Hernia. 2013 Apr;17(2):193-202.

[40] Yahchouchy-Chouillard E, Aura T, Picone O, Etienne JC, Fingerhut A. Incisional hernias. I. Related risk factors. Dig Surg. 2003;20(1):3-9.

[41] Maglinte, D.D., Gourtsoyiannis, N., Rex, D., Howard, T.J., Kelvin, F.M.: Classification of small bowel Crohn's subtypes based on multimodality imaging. Radiol. Clin. North Am. 41(2), 285–303 (2003)

[42] Bhatnagar, G., Stempel, C., Halligan, S., Taylor, S.A.: Utility of MR enterography and ultrasound for the investigation of small bowel CD. J. Magn. Reson. Imaging 45, 1573–1588 (2016)

[43] Lo Re, G., Midiri, M.: Crohn's Disease: Radiological Features and Clinical-Surgical Correlations. Springer, Heidelberg (2016)

[44] Gomoll´on, F., Dignass, A., Annese, V., Tilg, H., Van Assche, G., Lindsay, J.O., Peyrin-Biroulet, L., Cullen, G.J., Daperno, M., Kucharzik, T., et al.: 3rd European evidence-based consensus on the diagnosis and management of Crohn's disease 2016: part 1: diagnosis and medical management. J. Crohns Colitis 11, 3–25 (2016). jjw168

[45] Peloquin, J.M., Pardi, D.S., Sandborn, W.J., Fletcher, J.G., McCollough, C.H., Schueler, B.A., Kofler, J.A., Enders, F.T., Achenbach, S.J., Loftus, E.V.: Diagnostic ionizing radiation exposure in a population-based cohort of patients with inflammatory bowel disease. Am. J. Gastroenterol. 103(8), 2015–2022 (2008)

[46] Sinha, R., Verma, R., Verma, S., Rajesh, A.: Mr enterography of Crohn disease: part 1, rationale, technique, and pitfalls. Am. J. Roentgenol. 197(1), 76–79 (2011)

[47] Panes, J., Bouzas, R., Chaparro, M., Garc´ıa-S´anchez, V., Gisbert, J., Mart´ınez de Guere˜nu, B., Mendoza, J.L., Paredes, J.M., Quiroga, S., Ripoll´es, T., et al.: Systematic review: the use of ultrasonography, computed tomography and magnetic resonance imaging for the diagnosis, assessment of activity and abdominal complications of Crohn's disease. Aliment. Pharmacol. Ther. 34(2), 125–145 (2011)

[48] Steward, M.J., Punwani, S., Proctor, I., Adjei-Gyamfi, Y., Chatterjee, F., Bloom, S., Novelli, M., Halligan, S., Rodriguez-Justo, M., Taylor, S.A.: Non-perforating small bowel CD assessed by MRI enterography: derivation and histopathological validation of an MR-based activity index. Eur. J. Radiol. 81(9), 2080–2088 (2012)

[49] Lo Re, G., Cappello, M., Tudisca, C., Galia, M., Randazzo, C., Crax`ı, A.,

Camma, C., Giovagnoni, A., Midiri, M.: CT enterography as a powerful tool for the evaluation of inflammatory activity in Crohn's disease: relationship of CT findings with CDAI and acute-phase reactants. Radiol. Med. (Torino) 119(9), 658–666 (2014)

[50] Tolan, D.J., Greenhalgh, R., Zealley, I.A., Halligan, S., Taylor, S.A.: Mr enterographic manifestations of small bowel Crohn disease 1. Radiographics 30(2), 367–384 (2010)

[51] Sinha, R., Verma, R., Verma, S., Rajesh, A.: Mr enterography of Crohn disease: part 2, imaging and pathologic findings. Am. J. Roentgenol. 197(1), 80–85 (2011)

[52] Chaplot, S., Patnaik, L., Jagannathan, N.: Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomed. Signal Process. Control 1(1), 86–92 (2006)

[53] Cocosco, C.A., Zijdenbos, A.P., Evans, A.C.: A fully automatic and robust brain MRI tissue classification method. Med. Image Anal. 7(4), 513–527 (2003)

[54] Agnello, L., Comelli, A., Ardizzone, E., Vitabile, S.: Unsupervised tissue classification of brain MR images for voxel-based morphometry analysis. Int. J. Imaging Syst. Technol. 26(2), 136–150 (2016)

[55] Zhang, Y., Wu, L.: Weights optimization of neural network via improved BCO approach. Prog. Electromagnet. Res. 83, 185–198 (2008)

[56] Comelli, A., Agnello, L., Vitabile, S.: An ontology-based retrieval system for mammographic reports. In: 2015 IEEE Symposium on Computers and Communication (ISCC), pp. 1001–1006. IEEE (2015)

[57] Yeh, J.Y., Fu, J.: A hierarchical genetic algorithm for segmentation of multispectral human-brain MRI. Expert Syst. Appl. 34(2), 1285–1295 (2008)

[58] Patil, N., Shelokar, P., Jayaraman, V., Kulkarni, B.: Regression models using pattern search assisted least square support vector machines. Chem. Eng. Res. Des. 83(8), 1030–1037 (2005)

[59] Wang, F.F., Zhang, Y.R.: The support vector machine for dielectric target detection through a wall. Prog. Electromagnet. Res. Lett. 23, 119–128 (2011)

[60] Xu, Y., Guo, Y., Xia, L., Wu, Y.: An support vector regression based nonlinear modeling method for SiC MESFET. Prog. Electromagnet. Res. Lett. 2, 103–114 (2008)

[61] Li, D., Yang, W., Wang, S.: Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine. Comput. Electron. Agric. 74(2), 274–279 (2010)

[62] Son, Y.J., Kim, H.G., Kim, E.H., Choi, S., Lee, S.K.: Application of support vector machine for prediction of medication adherence in heart failure patients. Healthc. Inform. Res. 16(4), 253–259 (2010)

[63] Zhang, Y., Wang, S., Ji, G., Dong, Z.: An MR brain images classifier system

via particle swarm optimization and Kernel support vector machine. Sci. World J. 2013, 9 (2013)

[64] Tagluk, M.E., Akin, M., Sezgin, N.: Classıfication of sleep apnea by using wavelet transform and artificial neural networks. Expert Syst. Appl. 37(2), 1600–1607 (2010)

[65] Agnello, L., Comelli, A., Vitabile, S.: Feature dimensionality reduction for mammographic report classification. In: Pop, F., Ko_lodziej, J., Di Martino, B. (eds.) Resource Management for Big Data Platforms. CCN, pp. 311–337. Springer, Cham (2016). doi:10.1007/978-3-319-44881-7 15

[66] Martiskainen, P., J¨arvinen, M., Sk¨on, J.P., Tiirikainen, J., Kolehmainen, M., Mononen, J.: Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. Appl. Anim. Behav. Sci. 119(1), 32–38 (2009)

[67] Deris, A.M., Zain, A.M., Sallehuddin, R.: Overview of support vector machine in modeling machining performances. Procedia Eng. 24, 308–312 (2011)

[68] Bermejo, S., Monegal, B., Cabestany, J.: Fish age categorization from otolith images using multi-class support vector machines. Fish. Res. 84(2), 247–253 (2007)

[69] K. L. Newbold, M. Partridge, G. Cook, B. Sharma,P. Rhys-Evans, K. J. Harrington, and C. M. Nutting,"Evaluation of the role of 18fdg-pet/ct in radiotherapy target definition in patients with head and neck cancer," Acta Oncol. 47, 1229–1236 (2008).

[70] Niyazi M, Landrock S, Elsner A, et al. (2013) Automated biological target volume delineation for radiotherapy treatment planning using FDG-PET/CT. Radiat Oncol 8:180. doi: 10.1186/1748-717X-8-180

[71] Mi H, Petitjean C, Dubray B, et al. (2015) Robust feature selection to predict tumor treatment outcome. Artif Intell Med 64:195–204. doi: 10.1016/j.artmed.2015.07.002

[72] Alongi P, Caobelli F, Gentile R, et al. (2016) Recurrent bladder carcinoma: clinical and prognostic role of 18 F-FDG PET/CT. Eur J Nucl Med Mol Imaging. doi: 10.1007/s00259-016-3500-8

[73] Allegra E, Saita V, De Natale M, et al. (2017) Use of PET/CT to detect local and regional laryngeal cancer recurrence after surgery. Reports Med Imaging 10:31–36. doi: 10.2147/RMI.S124764

[74] Fletcher JW, Kinahan PE (2010) PET/CT Standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy. NIH Public Access 31:496–505. doi: 10.1053/j.sult.2010.10.001.PET/CT

[75] Larson SM, Erdi Y, Akhurst T, et al. (1999) Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG imaging. The visual response score and the change in total lesion glycolysis. Clin Positron Imaging 2:159–171. doi: 10.1016/S1095-0397(99)00016-3

[76] Stefano A, Porcino N, Banna G, et al. (2015) Metabolic response assessment in non-small cell lung cancer patients after platinum-based therapy: A preliminary analysis. Curr Med Imaging Rev 11:218–227.

[77] Soret M, Bacharach SL, Buvat II (2007) Partial-volume effect in PET tumor imaging. J Nucl Med 48:932–945. doi: 10.2967/jnumed.106.035774

[78] Hatt M, Laurent B, Ouahabi A, et al. (2018) The first MICCAI challenge on PET tumor segmentation. Med Image Anal 44:177–195. doi: 10.1016/j.media.2017.12.007

[79] Combi C, Oliboni B, Zardini A, Zerbato F (2016) Seamless Design of Decision-Intensive Care Pathways. Proc - 2016 IEEE Int Conf Healthc Informatics, ICHI 2016. doi: 10.1109/ICHI.2016.9

[80] Boellaard R, O'Doherty MJ, Weber WA, et al. (2010) FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: Version 1.0. Eur J Nucl Med Mol Imaging 37:181–200. doi: 10.1007/s00259-009-1297-4

[81] Foster B, Bagci U, Mansoor A, et al. (2014) A review on segmentation of positron emission tomography images. Comput Biol Med 50:76–96. doi: 10.1016/j.compbiomed.2014.04.014

[82] McGuinness K, O'Connor NE (2010) A comparative evaluation of interactive segmentation algorithms. Pattern Recognit 43:434–444. doi: 10.1016/j.patcog.2009.03.008

[83] Lankton S, Nain D, Yezzi A, Tannenbaum A (2007) Hybrid geodesic region-based curve evolutions for image segmentation. In: Hsieh J, Flynn MJ (eds). International Society for Optics and Photonics, p 65104U

[84] Stefano A, Vitabile S, Russo G, et al. (2016) A fully automatic method for biological target volume segmentation of brain metastases. Int J Imaging Syst Technol 26:29–37. doi: 10.1002/ima.22154

[85] Nariai T, Tanaka Y, Wakimoto H, et al. (2005) Usefulness of l -[methyl- 11 C] methionine—positron emission tomography as a biological monitoring tool in the treatment of glioma. J Neurosurg 103:498–507. doi: 10.3171/jns.2005.103.3.0498

[86] Stefano A, Vitabile S, Russo G, et al. (2017) An enhanced random walk algorithm for delineation of head and neck cancers in PET studies. Med Biol Eng Comput 55:897–908. doi: 10.1007/s11517-016-1571-0

[87] Prieto E, Lecumberri P, Pagola M, et al. (2012) Twelve automated thresholding methods for segmentation of PET images: a phantom study. Phys Med Biol 57:3963–3980. doi: 10.1088/0031-9155/57/12/3963

[88] Hofheinz F, Dittrich S, Pötzsch C, Hoff J van den (2010) Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold. Phys Med Biol 55:1099–113. doi: 10.1088/0031-9155/55/4/013

[89] Soffientini CD, De Bernardi E, Casati R, et al. (2017) Technical Note: A new zeolite PET phantom to test segmentation algorithms on heterogeneous activity

distributions featured with ground-truth contours. Med Phys. doi: 10.1002/mp.12014

[90] Feuardent J, Soret M, de Dreuille O, et al. (2003) Reliability of SUV estimates in FDG PET as a function of acquisition and processing protocols. Nucl Sci Symp Conf Rec 2003 IEEE 4:2877–2881 Vol.4. doi: 10.1109/NSSMIC.2003.1352484

[91] Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res. doi: 10.1613/jair.953

[92] Gonzalez-Reyna SE, Martinez-Trinidad JF, Carrasco-Ochoa JA, et al. (2014) Applying balancing techniques in traffic sign recognition. Artif Intell Res. doi: http://dx.doi.org/10.1016/j.oooo.2015.09.019

[93] Comelli A., Stefano A., Benfante V. RG (2018) Normal and Abnormal Tissue Classification in PET Oncological Studies. Pattern Recognit Image Anal 28:121–128. doi: 10.1134/S1054661818010054

[94] S. Armand, E. Watelain, E. Roux, M. Mercier, and F.-X. Lepoutre, "Linking clinical measurements and kinematic gait patterns of toe-walking using fuzzy decision trees," Gait Posture 25, 475–484 (2007).

[95] Dewalle-Vignion A-S, Yeni N, Petyt G, et al. (2012) Evaluation of PET volume segmentation methods. Nucl Med Commun 33:34–42. doi: 10.1097/MNM.0b013e32834d736f

[96] Day E, Betler J, Parda D, et al. (2009) A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients. Med Phys 36:4349–4358. doi: 10.1118/1.321309

[97] Chan TF, Vese LA (2001) Active contours without edges. IEEE Trans Image Process 10:266–277. doi: 10.1109/83.902291

[98] Comelli A, Stefano A, Russo G, et al. (2018) A smart and operator independent system to delineate tumours in Positron Emission Tomography scans. Comput Biol Med. doi: 10.1016/J.COMPBIOMED.2018.09.002

[99] Liu Y, Liu S, Nacif MS, et al. (2013) A framework to measure myocardial extracellular volume fraction using dual-phase low dose CT images. Med Phys. doi: 10.1118/1.4819936

[100] Hatt M, Lee JA, Schmidtlein CR, et al. (2017) Classi fi cation and evaluation strategies of auto-segmentation approaches for PET : Report of AAPM task group No.211. Med Phys 44

[101] Udupa JK, Leblanc VR, Zhuge Y, et al. (2006) A framework for evaluating image segmentation algorithms. Comput Med Imaging Graph 30:75–87. doi: 10.1016/j.compmedimag.2005.12.001

[102] Graham MM, Peterson LM, Hayward RM (2000) Comparison of simplified quantitative analyses of FDG uptake. Nucl Med Biol 27:647–655. doi: 10.1016/S0969-8051(00)00143-8

[103] Zhuang M, Dierckx RAJO, Zaidi H (2016) Generic and robust method for automatic segmentation of PET images using an active contour model. Med Phys 43:4483–4494. doi: 10.1118/1.4954844

[104] Foster B, Bagci U, Luna B, et al. (2013) Robust Segmentation and Accurate Target Definition for Positron Emission Tomography Images Using Affinity Propagation. 2013 IEEE 10TH Int. Symp. Biomed. IMAGING

[105] Hatt M, Tixier F, Pierce L, et al. (2017) Characterization of PET/CT images using texture analysis: the past, the present… any future? Eur J Nucl Med Mol Imaging. doi: 10.1007/s00259-016-3427-0

[106] Stefano A, Vitabile S, Russo G, et al. (2015) An automatic method for metabolic evaluation of gamma knife treatments. In: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). Springer, Cham, pp 579–589

[107] Comelli A, Stefano A, Russo G, et al (2019) K-nearest neighbor driving active contours to delineate biological tumor volumes. Eng Appl Artif Intell 81:133–144. doi: 10.1016/j.engappai.2019.02.005

[108] Comelli A, Stefano A, Bignardi S, et al (2019) Active contour algorithm with discriminant analysis for delineating tumors in positron emission tomography. Artif Intell Med 94:67–78. doi: 10.1016/J.ARTMED.2019.01.002

[109] Belhassen S, Zaidi H (2010) A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. Med Phys 37:1309–1324. doi: 10.1118/1.3301610

[110] Taha AA, Hanbury A (2015) Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 15:29. doi: 10.1186/s12880-015-0068-x

[111] Moghaddasi L, Bezak E, Marcu LG (2012) Current challenges in clinical target volume definition: Tumour margins and microscopic extensions. Acta Oncol (Madr) 51:984–995. doi: 10.3109/0284186X.2012.720381

[112] Stefano A, Gallivanone F, Messa CL, et al. (2014) Metabolic impact of Partial Volume Correction of [18F]FDG PET-CT oncological studies on the assessment of tumor response to treatment. Q J Nucl Med Mol Imaging 58:413–423.

[113] Berthon B, Spezi E, Galavis P, et al (2017) Toward a standard for the evaluation of PET-Auto-Segmentation methods following the recommendations of AAPM task group No. 211: Requirements and implementation. Med Phys. doi: 10.1002/mp.12312

[114] Stefano A, Vitabile S, Russo G, et al. (2013) A Graph-Based Method for PET Image Segmentation in Radiotherapy Planning: A Pilot Study. In: Petrosino A (ed) Image Anal. Process. Springer-Verlag Berlin, pp 711–720

[115] Sbei A, ElBedoui K, Barhoumi W, et al. (2017) Hybrid PET/MRI co-segmentation based on joint fuzzy connectedness and graph cut. Comput

Methods Programs Biomed 149:29–41. doi: 10.1016/j.cmpb.2017.07.006

[116] Lee JA (2010) Segmentation of positron emission tomography images: Some recommendations for target delineation in radiation oncology. Radiother Oncol 96:302–307. doi: 10.1016/j.radonc.2010.07.003

[117] Tan S, Li L, Choi W, et al. (2017) Adaptive region-growing with maximum curvature strategy for tumor segmentation in18F-FDG PET. Phys Med Biol. doi: 10.1088/1361-6560/aa6e20

[118] Hatt M, Cheze le Rest C, Descourt P, et al. (2010) Accurate Automatic Delineation of Heterogeneous Functional Volumes in Positron Emission Tomography for Oncology Applications. Int J Radiat Oncol Biol Phys 77:301–308. doi: 10.1016/j.ijrobp.2009.08.018

[119] Erickson BJ, Korfiatis P, Akkus Z, Kline TL (2017) Machine Learning for Medical Imaging. RadioGraphics. doi: 10.1148/rg.2017160130

[120] Li H, Thorstad WL, Biehl KJ, et al. (2008) A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours. Med Phys 35:3711–3721. doi: 10.1118/1.2956713

[121] Abdoli M, Dierckx RAJO, Zaidi H (2013) Contourlet-based active contour model for PET image segmentation. Med Phys. doi: 10.1118/1.4816296

[122] Onoma DP, Ruan S, Thureau S, et al. (2014) Segmentation of heterogeneous or small FDG PET positive tissue based on a 3D-locally adaptive random walk algorithm. Comput Med Imaging Graph 38:753–763. doi: 10.1016/j.compmedimag.2014.09.007

[123] Bi L, Kim J, Wen L, et al. (2015) Automated thresholded region classification using a robust feature selection method for PET-CT. In: 2015 IEEE 12th Int. Symp. Biomed. Imaging. IEEE, pp 1435–1438

[124] Pianykh OS (2004) Digital Imaging and Communications in Medicine: DICOM. Medicine (Baltimore). doi: 10.1007/978-3-540-74571-6

[125] Cover TM, Hart PE (1967) Nearest Neighbor Pattern Classification. IEEE Trans Inf Theory. doi: 10.1109/TIT.1967.1053964

[126] Zhou CY, Chen YQ (2006) Improving nearest neighbor classification with cam weighted distance. Pattern Recognit. doi: 10.1016/j.patcog.2005.09.004

[127] Comelli A, Terranova MC, Scopelliti L, et al (2018) A Kernel Support Vector Machine Based Technique for Crohn's Disease Classification in Human Patients. Springer, Cham, pp 262–273

[128] Son YJ, Kim HG, Kim EH, Choi S, Lee SK. Application of support vector machine for prediction of medication adherence in heart failure patients. Healthcare Informatics Research. 2010;16(4):253-259.

[129] Zhang Y, Wang S, Ji G, Dong Z. An MR brain images classifier system via particle swarm optimization and kernel support vector machine. The Scientific World Journal. 2013.

[130] Bermejo S, Monegal B, Cabestany J. Fish age categorization from otolith

images using multi-class support vector machines. Fisheries Research. 2007;84(2):247-253.

[131] Muniz A, Liu H, Lyons K, Pahwa R, Liu W, Nobre F, Nadal J. Comparison among probabilistic neural network, support vector machine and logistic regression for evaluating the effect of subthalamic stimulation in parkinson disease on ground re- action force during gait. Journal of biomechanics. 2010;43 (4):720-726.

[132] Mureşan M, Mureşan S, Bara T, Neagoe R, Sala D, Suciu B. Hernia recurrence long term follow-up after open procedures of abdominal wall plasty-prospective study including 142 patients. Cir Cir. 2016 Sep-Oct;84(5):376-83. doi: 10.1016/j.cir- cir.2015.10.012. Epub 2016 Jan 15.

[133] Murariu M, Bota N, Avram J. Causes of recurrent incisional herniation. Cercetãri Experimentale & Medico-Chirurgale. 2007:142-146.

[134] Hauters P, Desmet J, Gherardi D, Dewaele S, Poilvache H, Malvaux P. Assessment of predictive factors for recurrence in laparoscopic ventral hernia repair using a bridging technique. Surg Endosc. 2017 Sep;31(9):3656-3663. doi: 10.1007/s00464- 016-5401-0. Epub 2017 Jan 11.

[135] Rundo L, Stefano A, Militello C, et al. (2017) A fully automatic approach for multimodal PET and MR image segmentation in gamma knife treatment planning. Comput Methods Programs Biomed 144:77–96. doi: 10.1016/j.cmpb.2017.03.011

[136] Aspert N, Santa-Cruz D, Ebrahimi T (2002) MESH: Measuring errors between surfaces using the Hausdorff distance. Proc - 2002 IEEE Int Conf Multimed Expo, ICME 2002. doi: 10.1109/ICME.2002.103587

[137] Gallivanone F, Stefano A, Grosso E, et al. (2011) PVE correction in PET-CT whole-body oncological studies from PVE-affected images images. IEEE Trans Nucl Sci 58:736–747.

[138] Berthon B, Marshall C, Edwards a, et al. (2013) Influence of cold walls on PET image quantification and volume segmentation: a phantom study. Med Phys 40:082505. doi: 10.1118/1.4813302

[139] R. Boellaard, R. Delgado-Bolton, W. J. G. Oyen, F. Giammarile, K. Tatsch, W. Eschner, F. J. Verzijlbergen, S. F. Barrington, L. C. Pike, W. A. Weber, et al., "Fdg pet/ct: Eanm procedure guidelines for tumour imaging: version 2.0," Europ. J. Nucl. Med. Mol. Imag. 42, 328–354 (2015).

[140] F. Gallivanone, A. Stefano, E. Grosso, C. Canevari, L. Gianolli, C. Messa, M. C. Gilardi, and I. Castiglioni, "Pve correction in pet-ct wholebody oncological studies from pve-affected images," IEEE Trans. Nucl. Sci. 58, 736–747 (2011).

[141] Wang X-Y, Zhao Y-F, Liu Y, et al. (2017) Comparison of different automated lesion delineation methods for metabolic tumor volume of 18F-FDG PET/CT in patients with stage I lung adenocarcinoma. Medicine (Baltimore). doi: 10.1097/MD.0000000000009365

[142] Belli ML, Mori M, Broggi S, et al. (2018) Quantifying the robustness of [18F]FDG-PET/CT radiomic features with respect to tumor delineation in head and neck and pancreatic cancer patients. Phys Medica 49:105–111. doi: 10.1016/j.ejmp.2018.05.013

[143] Licari L, Salamone G, Campanella S, et al Use of the KSVM-based system for the definition, validation and identification of the incisional hernia recurrence risk factors. G Chir 40:32–38.

[144] Göçeri E (2016) Fully automated liver segmentation using Sobolev gradient-based level set evolution. Int j numer method biomed eng. doi: 10.1002/cnm.2765

[145] Goceri E, Dura E (2013) A Level Set Method with Sobolev Gradient and Haralick Edge Detection. 4th World Conf.on Inf Technol (WCIT 2013) 5:131–140.

[146] Comelli A, Bruno A, Di Vittorio ML, et al (2017) Automatic Multi-seed Detection for MR Breast Image Segmentation. Springer, Cham, pp 706–717